```
                                                                  sau96I
                                                                  haeIII
                                                                  asuI
                                                                 sau96I
                                                                 nlaIV
                                                                 hgiJII
                                                                 eco□109I
                                                                 bsp1286
                                                                 banII
                                                                 asuI
                                                                nlaIV
                        mnlI                                 aval apaI mnlI
                     mboII                                mnlI  eco□109I     tth111I
  1     GGA CTT GTC TTC CTC GTC CTG CTG TTC CTC GGG GCC CTC GGA CTG
-18     Gly Leu Val Phe Leu Val Leu Leu Phe Leu Gly Ala Leu Gly Leu

                  haeIII
                 eaeI                                              hinPI
                 cfrI                                              hhaI
 46 TGT CTG GCT GGC CGT AGG AGA AGG AGT GTT CAG TGG TGC GCC GTA TCC
 -3 Cys Leu Ala Gly Arg Arg Arg Arg Ser Val Gln Trp Cys Ala Val Ser

                haeIII
               mnlI
            aval haeI
 94 CAA CCC GAG GCC ACA AAA TGC TTC CAA TGG CAA AGG AAT ATG AGA AAA
 14 Gln Pro Glu Ala Thr Lys Cys Phe Gln Trp Gln Arg Asn Met Arg Lys

                   mnlI          fnu4HI
               sau96I            bbvI                 pleI
               haeIII           aluI                  hinfI          bsrI
               asuI           pvuII                 bsmaI         fokI
142 GTG CTG GGC CCT CCT GTC AGC TGC ATA AAG AGA GAC TCC CCC ATC CAG
 30 Val Arg Gly Pro Pro Val Ser Cys Ile Lys Arg Asp Ser Pro Ile Gln

                haeIII
               haeI
            scrFI                          haeIII
            ecoRII                        sau96I
            bstNI                         asuI    sfaNI
190 TGT ATC CAG GCC ATT GCG GAA AAC AGG GCC GAT GCT GTG ACC CTT GAT
 46 Cys Ile Gln Ala Ile Ala Glu Asn Arg Ala Asp Ala Val Thr Leu Asp

                                          sau96I
                                          nlaIV
                                      scrFI
                                      ecoRII
                                      bstNI
                                    haeIII
                                 stuI    haeIII
                           mnlI  haeI    asuI
238 GGT GGT TTC ATA TAC GAG GCA GGC CTG GCC CCC TAC AAA CTG CGA CCT
 62 Gly Gly Phe Ile Tyr Glu Ala Gly Leu Ala Pro Tyr Lys Leu Arg Pro
```

FIG.-1A

```
                                  sau96I
                                  avaII
                                  asuI
          fnu4HI        accI    nlaIV
286 GTA GCG GCG GAA GTC TAC GGG ACC GAA AGA CAG CCA CGA ACT CAC TAT
 78 Val Ala Ala Glu Val Tyr Gly Thr Glu Arg Gln Pro Arg Thr His Tyr

                                                 fnu4HI
                                 mboII           bbvI        alul
                        hphI                  fnu4HI alul    pvuII
334 TAT GCC GTG GCT GTG GTG AAG AAG GGC GGC AGC TTT CAG CTG AAC GAA
 94 Tyr Ala Val Ala Val Val Lys Lys Gly Gly Ser Phe Gln Leu Asn Glu

                                              haeIII       sau96I
                                            stuI           avaII
                                    bglI    haeI           asuI       fokI
382 CTG CAA GGT CTG AAG TCC TGC CAC ACA GGC CTT CGC AGG ACC GCT GGA
110 Leu Gln Gly Leu Lys Ser Cys His Thr Gly Leu Arg Arg Thr Ala Gly

                                                                    sau961
                                                                    avaII
                                                                    asuI
                                                                  nlaIV
430 TGG AAT GTC CCT ACA GGG ACA CTT CGT CCA TTC TTG AAT TGG ACG GGT
126 Trp Asn Val Pro Thr Gly Thr Leu Arg Pro Phe Leu Asn Trp Thr Gly

            hgiJII            alul
            bsp1286         fnu4HI
            banII           bbvI                    ddeI      alul
         ddeI             mnlI pvuII           mboII        pvuII
478 CCA CCT GAG CCC ATT GAG GCA GCT GTG CAG TTC TTC TCA GCC AGC TGT
142 Pro Pro Glu Pro Ile Glu Ala Ala Val Gln Phe Phe Ser Ala Ser Cys

          mspI
          hpaII
         scrFI
         ncil
         cauII
526 GTT CCC GGT GCA GAT AAA GGA CAG TTC CCC AAC CTG TGT CGC CTG TGT
158 Val Pro Gln Ala Asp Lys Gly Gln Phe Pro Asn Leu Cys Arg Leu Cys

                                                      nlaIV
                                                   scrFI
                                                   ecoRII
                                                mnlI bstNI      rsaI
574 GCG GGG ACA GGG GAA AAC AAA TGT GCC TTC TCC TCC CAG GAA CCG TAC
174 Ala Gly Thr Gly Glu Asn Lys Cys Ala Phe Ser Ser Gln Glu Pro Tyr
```

FIG.-1B

```
                    nlaIV
                    hgiCI
          aluI      banI                ddeI  bsmaI          bsmaI
622 TTC AGC TAC TCT GGT GCC TTC AAG TGT CTG AGA GAC GGG GCT GGA GAC
190 Phe Ser Tyr Ser Gly Ala Phe Lys Cys Leu Arg Asp Gly Ala Gly Asp

                                              sau96I
                                              avaII
                                              asuI
                                             ppuMI
                          hgiAI               eco0109I
                          bsp1286           mnlI                mnlI
670 GTG GCT TTT ATC AGA GAG AGC ACA GTG TTT GAG GAC CTG TCA GAC GAG
206 Val Ala Phe Ile Arg Glu Ser Thr Val Phe Glu Asp Leu Ser Asp Glu

718 GCT GAA AGG GAC GAG TAT GAG TTA CTC TGC CCA GAC AAC ACT CGG AAG
222 Ala Glu Arg Asp Glu Tyr Glu Leu Leu Cys Pro Asp Asn Thr Arg Lys

                                              scrFI
                                              nciI
                                              mspI
                                              hpaII
                                              cauII
                                             xmaI sau96I
                                             smaI nlaIV
                                             scrFI
                                             nciI avaII
                                             cauII
                                             aval asuI
                                          sau96I ppuMI
                                          hae III nlaIV
    bsrI                                  asuI   eco0109I       nlaIII
766 CCA GTG GAC AAG TTC AAA GAC TGC CAT CTG GCC CGG GTC CCT TCT CAT
238 Pro Val Asp Lys Phe Lys Asp Cys His Leu Ala Arg Val Pro Ser His

                                                sfaNI
                                               fokI                mboII
    bglI          draIII                     mnlI                hinfI
814 GCC GTT GTG GCA CGA AGT GTG AAT GGC AAG GAG GAT GCC ATC TGG AAT
254 Ala Val Val Ala Arg Ser Val Asn Gly Lys Glu Asp Ala Ile Trp Asn

                scrFI
                ecoRII
                bstNI                                      hphI
862 CTT CTC CGC CAG GCA CAG GAA AAG TTT GGA AAG GAC AAG TCA CCG AAA
270 Leu Leu Arg Gln Ala Gln Glu Lys Phe Gly Lys Asp Lys Ser Pro Lys
```

FIG.-1C

```
                                                        sau3AI
                                                        mboI
                                                        dpnI
                                                      xhoII
          aluI                                        bstYI
        bstXI          nlaIV                          bglII
910 TTC CAG CTC TTT GGC TCC CCT AGT GGG CAG AAA GAT CTG CTG TTC AAG
286 Phe Gln Leu Phe Gly Ser Pro Ser Gly Gln Lys Asp Leu Leu Phe Lys

                                    nlaIV
                                    hgiCI
     pleI                         mnlI  bsp1286   mnlI
     hinfI                      taqI  banI    avaI          hinfI
958 GAC TCT GCC ATT GGG TTT TCG AGG GTG CCC CCG AGG ATA GAT TCT GGG
302 Asp Ser Ala Ile Gly Phe Ser Arg Val Pro Pro Arg Ile Asp Ser Gly

                     mspI
          styI       hpaII
       rsaI      nlaIV                        fokI            mnlI
1006 CTG TAC CTT GGC TCC GGC TAC TTC ACT GCC ATC CAG AAC TTG AGG AAA
318 Leu Tyr Leu Gly Ser Gly Tyr Phe Thr Ala Ile Gln Asn Leu Arg Lys

                                    mspI
                                    hpaII      thaI
                                   scrFI       fnuDII
                                   nciI        bstUI
              mnlI          fnu4HI           hinPI
          mnlI              bbvI cauII       hhaI
1054 AGT GAG GAG GAA GTG GCT GCC CGG CGT GCG CGG GTC GTG TGG TGT GCG
344 Ser Glu Glu Glu Val Ala Ala Arg Arg Ala Arg Val Val Trp Cys Ala

                                  hinPI
                                 mstI
                                 fspI
                               fnu4HI
                               aluI hhaI            bstXI
                        alwNI bbvI                  bsrI
1102 GTG GGC GAG CAG GAG CTG CGC AAG TGT AAC CAG TGG AGT GGC TTG AGC
350 Val Gly Glu Gln Glu Leu Arg Lys Cys Asn Gln Trp Ser Gly Leu Ser

          fnu4HI                          mnlI
          bbvI        bspMI    mnlI haeIII              mnlI     sfaNI
1150 GAA GGC AGC GTG ACC TGC TCC TCG GCC TCC ACC ACA GAG GAC TGC ATC
366 Glu Gly Ser Val Thr Cys Ser Ser Ala Ser Thr Thr Glu Asp Cys Ile

        scrFI
        ecoRII                                    bstXI
        bstNI                          aluI  sfaNI nlaIII      fokI   mnlI
1198 GCC CTG GTG CTG AAA GGA GAA GCT GAT GCC ATG AGT TTG GAT GGA GGA
382 Ala Leu Val Leu Lys Gly Glu Ala Asp Ala Met Ser Leu Asp Gly Gly
```

FIG.-1D

```
                          nlaIII                nlaIV       scrFI
                          sphI                  hgiCI       ecoRII
                rsaI      nspCIx                banI        bstNI
1246 TAT GTG TAC ACT GCA TGC AAA TGT GGT TTG GTG CCT GTC CTG GCA GAG
 398 Tyr Val Tyr Thr Ala Cys Lys Cys Gly Leu Val Pro Val Leu Ala Glu

                                                sau3AI
                                                mboI
                                                dpnI
                                                alwI
1294 AAC TAC AAA TCC CAA CAA AGC AGT GAC CCT GAT CCT AAC TGT GTG GAT
 414 Asn Tyr Lys Ser Gln Gln Ser Ser Asp Pro Asp Pro Asn Cys Val Asp

                                                              sau3AI
                                                              mboI
           ecoNI          ecoRV                               dpnI
1342 AGA CCT GTG GAA GGA TAT CTT GCT GTG GCG GTG GTT AGG AGA TCA GAC
 430 Arg Pro Val Glu Gly Tyr Leu Ala Val Ala Val Val Arg Arg Ser Asp

                    scrFI
                    ecoRII
                    bstNI
1390 ACT AGC CTT ACC TGG AAC TCT GTG AAA GGC AAG AAG TCC TGC CAC ACC
 446 Thr Ser Leu Thr Trp Asn Ser Val Lys Gly Lys Lys Ser Cys His Thr

                                                      haeIII
                                                nlaIII
                                               styI sau96I      mboII
                          pstI                 ncoI asuI       earI
1438 GCC GTG GAC AGG ACT GCA GGC TGG AAT ATC CCC ATG GGC CTG CTC TTC
 462 Ala Val Asp Arg Thr Ala Gly TrP Asn Ile Pro Met Gly Leu Leu Phe

                  nlaIV
                 hgiJII
                 bsp1286
                 banII                              sspI                alwI bsp1286
1486 AAC CAG ACG GGC TCC TGC AAA TTT GAT GAA TAT TTC AGT CAA AGC TGT
 478 Asn Gln Thr Gly Ser Cys Lys Phe Asp Glu Tyr Phe Ser Gln Ser Cys

                                 sau3AI
                                 mboI
                                 Dpn I
           scrFI                 xhoII
           ecoRII                bstYI             hgiAI
           bstNI          aval   bglII             bsp1286
1534 GCC CCT GGG TCT GAC CCG AGA TCT AAT CTC TGT GCT CTG TGT ATT GGC
 484 Ala Pro Gly Ser Asp Pro Arg Ser Asn Leu Cys Ala Leu Cys Ile Gly
```

FIG.-1E

```
                  hphI                  bsp1286
1582 GAC GAG CAG GGT GAG AAT AAG TGC GTG CCC AAC AGC AAT GAG AGA TAC
 510 Asp Glu Gln Gly Glu Asn Lys Cys Val Pro Asn Ser Asn Glu Arg Tyr

                                    nlaIV
                                    hgiCI
                                    banI scrFI
                                 mspI    ecoRII
                  bsrI           hpaII   bstNI  ddeI bsmI       bsmaI
1630 TAC GGC TAC ACT GGG GCT TTC CGG TGC CTG GCT GAG AAT GCT GGA GAC
 526 Tyr Gly Tyr Thr Gly Ala Phe Arg Cys Leu Ala Glu Asn Ala Gly Asp

1678 GTT GCA TTT GTG AAA GAT GTC ACT GTC TTG CAG AAC ACT GAT GGA AAT
 542 Val Ala Phe Val Lys Asp Val Thr Val Leu Gln Asn Thr Asp Gly Asn

                                                                fnu4HI
                                                                bbvI
                                                             hinPI
             mnlI nlaIII  ddeI              aluI                hhaI
1726 AAC AAT GAG GCA TGG GCT AAG GAT TTG AAG CTG GCA GAC TTT GCG CTG
 558 Asn Asn Glu Ala Trp Ala Lys Asp Leu Lys Leu Ala Asp Phe Ala Leu

              taqI                                              fnu4HI
           mnlI                                    mnlI         bbvI
          bglI                                   ddeI          aluI
1774 CTG TGC CTC GAT GGC AAA CGG AAG CCT GTG ACT GAG GCT AGA AGC TGC
 574 Leu Cys Leu Asp Gly Lys Arg Lys Pro Val Thr Glu Ala Arg Ser Cys

                  sau96I
                  nlaIV
               nlaIII
              styI haeIII
              ncoI asuI    hinfI nlaIII        bsmaI  fokI
1822 CAT CTT GCC ATG GCC CCG AAT CAT GCC GTG GTG TCT CGG ATG GAT AAG
 590 His Leu Ala Met Ala Pro Asn His Ala Val Val Ser Arg Met Asp Lys

                              fnu4HI
               ecoNI      alwNI bbvI
1870 GTG GAA CGC CTG AAA CAG GTG CTG CTC CAC CAA CAG GCT AAA TTT GGG
 606 Val Glu Arg Leu Lys Gln Val Leu Leu His Gln Gln Ala Lys Phe Gly

              sau3AI
              mboI           mspI
              dpnI           hpaII
             xhoII         scrFI
             bstYI         nciI
             alwI          cauII                        bsrI
1919 AGA AAT GGA TCT GAC TGC CCG GAC AAG TTT TGC TTA TTC CAG TCT GAA
 622 Arg Asn Gly Ser Asp Cys Pro Asp Lys Phe Cys Leu Phe Gln Ser Glu
```

FIG.-1F

```
                                                                  haeIII
                                                                  haeI
                                                                  eaeI          styI
                                                  ddeI            cfrI     pleI ncoI
                                                 draIII           ballI    hinfI
1966 ACC AAA AAC CTT CTG TTC AAT GAC AAC ACT GAG TGT CTG GCC AGA CTC
 638 Thr Lys Asn Leu Leu Phe Asn Asp Asn Thr Glu Cys Leu Ala Arg Leu

                                                  sau96I
                                                  avaII
                                                  asuI
     nlaIII           ndeI        sspI      nlaIV
2014 CAT GGC AAA ACA ACA TAT GAA AAA TAT TTG GGA CCA CAG TAT GTC GCA
 654 His Gly Lys Thr Thr Tyr Glu Lys Tyr Leu Gly Pro Gln Tyr Val Ala

                                                                  scrFI
                                                                  ecoRII
                                  hgiAI                           bstNI
                                  bsp1286    mnlI      mnlI
2062 GGC ATT ACT AAT CGT AAA AAG TGC TCA ACC TCC CCC CTC CTG GAA GCC
 670 Gly Ile Thr Asn Leu Lys Lys Cys Ser Thr Ser Pro Leu Leu Glu Ala

                 ddeI
               mstII
               mnlI                                  sau96I
               eco81I                          mboII haeIII
         ecoRI bsu36I                         mboII  asuI aluI
2110 TGT GAA TTC CTC AGG AAG TAA  AACCGAAGAA GATGGCCCAG CTCCCCAAGA
 685 Cys Glu Phe Leu Arg Lys OC*

                                                                    styI
                                                                  haeIII
                                                                  sau96I
                                                 mboII    scrFI     asuI
        ddeI                                     earI     ecoRII  nlaIV
       mnlI                                     aluI      bstNI   eco0109I nlaIV
2161 AAGCCTCAGC CATTCACTGC CCCCAGCTCT TCTCCCCAGG TGTGTTGGGG CCTTGGCTCC

     ecoNI                    fokI                                    ddeI
2221 CCTGCTGAAG GTGGGGATTG CCCATCCATC TGCTTACAAT TCCCTGCTGT CGTCTTAGCA

2281 AGAAGTAAAA TGAGAAATTT TGTTGATATT CAAAAAAAA

>LENGTH: 2319
```

FIG.-1G

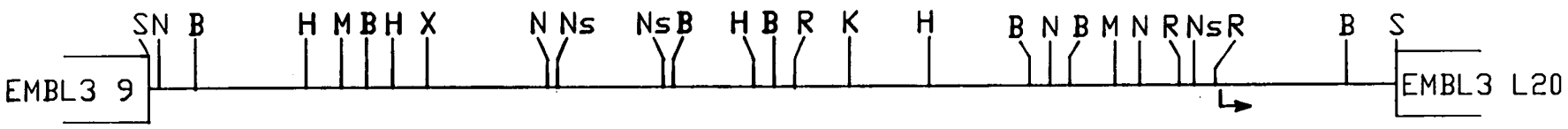
```
  1 GACTCCTAGG GGCTTGCAGA CCTAGTGGGA GAGAAAGAAC ATCGCAGCAG CCAGGCAGAA CCAGGACAGG TGAGGTGCAG GCTGGCTTTC CTCTCGCAGC
101 GCGGTGTGGA GTCCTGTCCT GCCTCAGGGC TTTTCGGAGC CTGGATCCTC AAGGAACAAG TAGACCTGGC CGCGGGGAGT GGGGAGGGAA GGGGTGTCTA
201 TTGGGCAACA GGGCGGCAAA GCCCTGAATA AAGGGGCGCA GGGCAGGCGC AAGTGCAGAG CCTTCGTTTG CCAAGTCGCC TCCAGACCGC AGAC ATG AAA CTT
-19                                                                                                        M   K   L
304 GTC TTC CTC GTC CTG CTG TTC CTC GGG GCC CTC GGA CTG TGT CTG GCT GGC CGT AGG AGA AGG AGT GTT CAG TGG TGC GCC GTA TCC
-16 V   F   L   V   L   L   F   L   G   A   L   G   L   C   L   A   G   R   R   R   R   S   V   Q   W   C   A   V   S
391 CAA CCC GAG GCC ACA AAA TGC TTC CAA TGG CAA AGG AAT ATG AGA AAA GTG CGT GGC CCT CCT GTC AGC TGC ATA AAG AGA GAC TCC
 14 Q   P   E   A   T   K   C   F   Q   W   Q   R   N   M   R   K   V   R   G   P   P   V   S   C   I   K   R   D   S
478 CCC ATC CAG TGT ATC CAG GCC ATT GCG GAA AAC AGG GCC GAT GCT GTG ACC CTT GAT GGT GGT TTC ATA TAC GAG GCA GGC CTG GCC
 43 P   I   Q   C   I   Q   A   I   A   E   N   R   A   D   A   V   T   L   D   G   G   F   I   Y   E   A   G   L   A
565 CCC TAC AAA CTG CGA CCT GTA GCG GCG GAA GTC TAC GGG ACC GAA AGA CAG CCA CGA ACT CAC TAT TAT GCC GTG GCT GTG GTG AAG
 72 P   Y   K   L   R   P   V   A   A   E   V   Y   G   T   E   R   Q   P   R   T   H   Y   Y   A   V   A   V   V   K
652 AAG GGC GGC AGC TTT CAG CTG AAC GAA CTG CAA GGT CTG AAG TCC TGC CAC ACA GGC CTT CGC AGG ACC GCT GGA TGG AAT GTC CCT
101 K   G   G   S   F   Q   L   N   E   L   Q   G   L   K   S   C   H   T   G   L   R   R   T   A   G   W   N   V   P
739 ACA GGG ACA CTT CGT CCA TTC TTG AAT TGG ACG GGT CCA CCT GAG CCC ATT GAG GCA GCT GTG GCC AGG TTC TTC TCA GCC AGC TGT
130 T   G   T   L   R   P   F   L   N   W   T   G   P   P   E   P   I   E   A   A   V   A   R   F   F   S   A   S   C
826 GTT CCC GGT GCA GAT AAA GGA CAG TTC CCC AAC CTG TGT CGC CTG TGT GCG GGG ACA GGG GAA AAC AAA TGT GCC TTC TCC TCC CAG
159 V   P   G   A   D   K   G   Q   F   P   N   L   C   R   L   C   A   G   T   G   E   N   K   C   A   F   S   S   Q
913 GAA CCG TAC TTC AGC TAC TCT GGT GCC TTC AAG TGT CTG AGA GAC GGG GCT GGA GAC GTG GCT TTT ATC AGA GAG AGC ACA GTG TTT
188 E   P   Y   F   S   Y   S   G   A   F   K   C   L   R   D   G   A   G   D   V   A   F   I   R   E   S   T   V   F
1000 GAG GAC CTG TCA GAC GAG GCT GAA AGG GAC GAG TAT GAG TTA CTC TGC CCA GAC AAC ACT CGG AAG CCA GTG GAC AAG TTC AAA GAC
217 E   D   L   S   D   E   A   E   R   D   E   Y   E   L   L   C   P   D   N   T   R   K   P   V   D   K   F   K   D
1087 TGC CAT CTG GCC CGG GTC CCT TCT CAT GCC GTT GTG GCA CGA AGT GTG AAT GGC AAG GAG GAT GCC ATC TGG AAT CTT CTC CGC CAG
246 C   H   L   A   R   V   P   S   H   A   V   V   A   R   S   V   N   G   K   E   D   A   I   W   N   L   L   R   Q
1174 GCA CAG GAA AAG TTT GGA AAG GAC AAG TCA CCG AAA TTC CAG CTC TTT GGC TCC CCT AGT GGG CAG AAA GAT CTG CTG TTC AAG GAC
275 A   Q   E   K   F   G   K   D   K   S   P   K   F   Q   L   F   G   S   P   S   G   Q   K   D   L   L   F   K   D
1261 TCT GCC ATT GGG TTT TCG AGG GTG CCC CCG AGG ATA GAT TCT GGG CTG TAC CTT GGC TCC GGC TAC TTC ACT GCC ATC CAG AAC TTG
304 S   A   I   G   F   S   R   V   P   P   R   I   D   S   G   L   Y   L   G   S   G   Y   F   T   A   I   Q   N   L
1348 AGG AAA AGT GAG GAG GAA GTG GCT GCC CGG CGT GCG CGG GTC GTG TGG TGT GCG GTG GGC GAG CAG GAG CTG CGC AAG TGT AAC CAG
333 R   K   S   E   E   E   V   A   A   R   R   A   R   V   V   W   C   A   V   G   E   Q   E   L   R   K   C   N   Q
```

FIG.-2A

```
1435 TGG AGT GGC TTG AGC GAA GGC AGC GTG ACC TGC TCC TCG GCC TCC ACC ACA GAG GAC TGC ATC GCC CTG GTG CTG AAA GGA GAA GCT
 362 W   S   G   L   S   E   G   S   V   T   C   S   S   A   S   T   T   E   D   C   I   A   L   V   L   K   G   E   A
1522 GAT GCC ATG AGT TTG GAT GGA GGA TAT GTG TAC ACT GCA TGC AAA TGT GGT TTG GTG CCT GTC CTG GCA GAG AAC TAC AAA TCC CAA
 391 D   A   M   S   L   D   G   G   Y   V   Y   T   A   C   K   C   G   L   V   P   V   L   A   E   N   Y   K   S   Q
1609 CAA AGC AGT GAC CCT GAT CCT AAC TGT GTG GAT AGA CCT GTG GAA GGA TAT CTT GCT GTG GCG GTG GTT AGG AGA TCA GAC ACT AGC
 420 Q   S   S   D   P   D   P   N   C   V   D   R   P   V   E   G   Y   L   A   V   A   V   V   R   R   S   D   T   S
1696 CTT ACC TGG AAC TCT GTG AAA GGC AAG AAG TCC TGC CAC ACC GCC GTG GAC AGG ACT GCA GGC TGG AAT ATC CCC ATG GGC CTG CTC
 449 L   T   W   N   S   V   K   G   K   K   S   C   N   T   A   V   D   R   T   A   G   W   N   I   P   M   Q   L   L
1783 TTC AAC CAG ACG GGC TCC TGC AAA TTT GAT GAA TAT TTC AGT CAA AGC TGT GCC CCT GGG TCT GAC CCG AGA TCT AAT CTC TGT GCT
 478 F   N   Q   T   G   S   C   K   F   D   E   Y   F   S   Q   S   C   A   P   G   S   D   P   R   S   N   L   C   A
1870 CTG TGT ATT GGC GAC GAG CAG GGT GAG AAT AAG TGC GTG CCC AAC AGC AAC GAG AGA TAC TAC GGC TAC ACT GGG GCT TTC CGG TGC
 507 L   C   I   G   D   E   Q   G   E   N   K   C   V   P   N   S   N   E   R   Y   Y   G   Y   T   G   A   F   R   C
1957 CTG GCT GAG AAT GCT GGA GAC GTT GCA TTT GTG AAA GAT GTC ACT GTC TTG CAG AAC ACT GAT GGA AAT AAC AAT GAG GCA TGG GCT
 536 L   A   E   N   A   G   D   V   A   F   V   K   D   V   T   V   L   Q   N   T   D   G   N   N   N   E   A   W   A
2044 AAG GAT TTG AAC CTG GCA GAC TTT GCG CTG CTG TGC CTC GAT GGC AAA CGG AAG CCT GTG ACT GAC GCT AGA AGC TGC CAT CTT GCC
 565 K   D   L   K   L   A   D   F   A   L   L   C   L   D   G   K   R   K   P   V   T   E   A   R   S   C   H   L   A
2131 ATG GCC CCG AAT CAT GCC GTG GTG TCT CGG ATG GAT AAG GTG GAA CGC CTG AAA CAG GTG CTG CTC CAC CAA CAG GCT AAA TTT GGG
 594 M   A   P   N   H   A   V   V   S   R   M   D   K   V   E   R   L   K   Q   V   L   L   H   Q   Q   A   K   F   G
2218 AGA AAT GGA TCT GAC TGC CCG GAC AAG TTT TGC TTA TTC CAG TCT GAA ACC AAA AAC CTT CTG TTC AAT GAC AAC ACT GAG TGT CTG
 623 R   N   G   S   D   C   P   D   K   F   C   L   F   Q   S   E   T   K   N   L   L   F   N   D   N   T   E   C   L
2305 GCC AGA CTC CAT GGC AAA ACA ACA TAT GAA AAA TAT TTG GGA CCA CAG TAT GTC GCA GGC ATT ACT AAT CTG AAA AAG TGC TCA ACC
 652 A   R   L   H   G   K   T   T   Y   E   K   Y   L   G   P   Q   Y   V   A   G   I   T   N   L   K   K   C   S   T
2392 TCC CCC CTC CTG GAA GCC TGT GAA TTC CTC AGG AAG TAA  AACCGAAGAA GATGGCCCAG CTCCCCAAGA AAGCCTCAGC CATTCACTGC CCCCAGCTCT
 681 S   P   L   L   E   A   C   E   F   L   R   K   O
2491 TCTCCCCAGG TGTGTTGGGG CCTTGGCTCC CCTGCTGAAG GTGGGGATTG CCCATCCATC TGCTTACAAT TCCCTGCTGT CGTCTTAGCA AGAAGTAAAA
2591 TGAGAAATTT TGTTGATATT CAAAAAAAA
```

FIG.-2B

RESTRICTION MAP OF 5' - FLANKING REGION OF αS1 CASEIN GENE



1Kb

S - SalI

B - BalII

R - EcoRI

H - HindIII

M - SmaI

N - NcoI

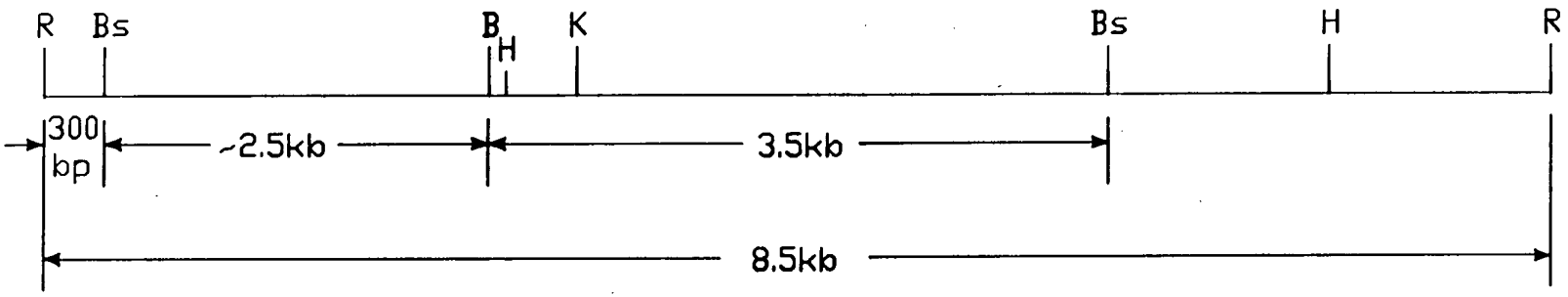
Ns - NsiI

X - XhoI

K - KpnI

FIG.-3

RESTRICTION MAP OF 3' FLANKING
REGION OF α S1 CASEIN GENE



1kb

R - EcoRI

Bs - Bst EII

B - BamHI

H - HindIII

K - KpnI

FIG.-4

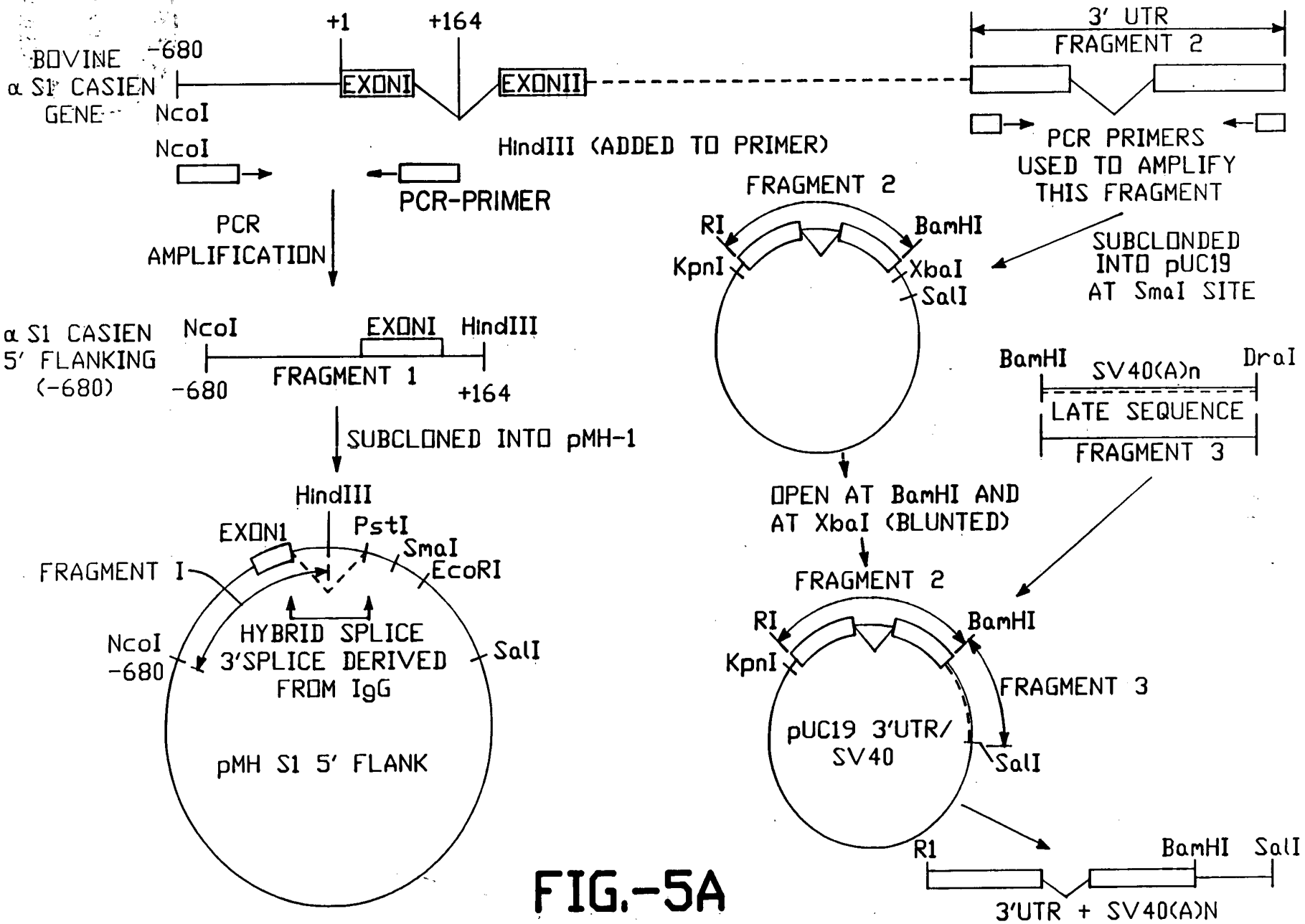

BOVINE α S1 CASIEN GENE
-680
+1
+164
EXONI
EXONII
3' UTR
FRAGMENT 2
NcoI
NcoI
HindIII (ADDED TO PRIMER)
PCR-PRIMER
PCR PRIMERS USED TO AMPLIFY THIS FRAGMENT
PCR AMPLIFICATION
α S1 CASIEN 5' FLANKING (-680)
NcoI
EXONI
HindIII
-680
FRAGMENT 1
+164
SUBCLONED INTO pMH-1
FRAGMENT 2
RI
BamHI
KpnI
XbaI
SalI
SUBCLONED INTO pUC19 AT SmaI SITE
BamHI
SV40(A)n
DraI
LATE SEQUENCE
FRAGMENT 3
OPEN AT BamHI AND AT XbaI (BLUNTED)
HindIII
EXON1
PstI
SmaI
EcoRI
FRAGMENT I
NcoI
-680
HYBRID SPLICE
3'SPLICE DERIVED FROM IgG
SalI
pMH S1 5' FLANK
FRAGMENT 2
RI
BamHI
KpnI
FRAGMENT 3
pUC19 3'UTR/ SV40
SalI
R1
BamHI
SalI
3'UTR + SV40(A)N

FIG.-5A

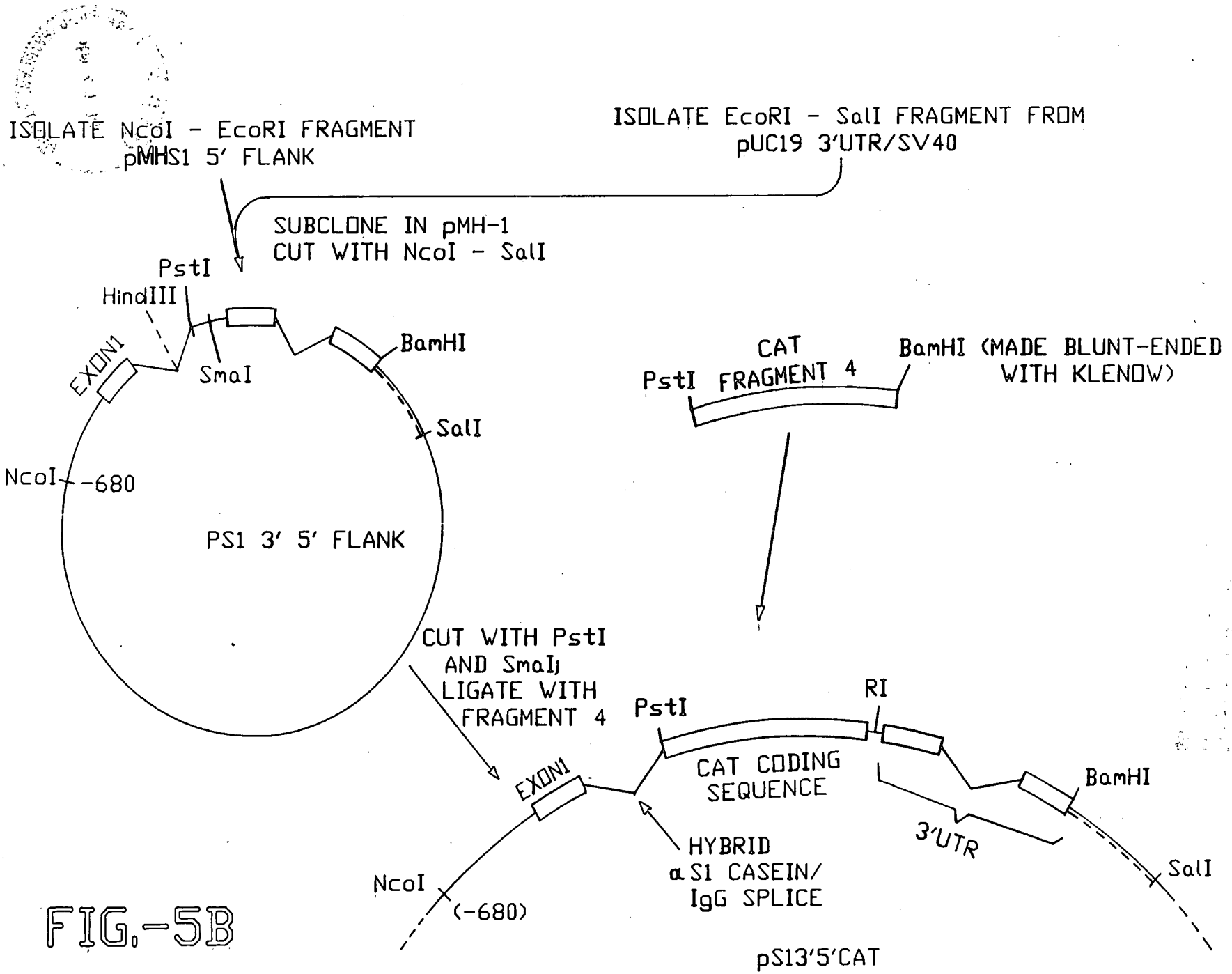

ISOLATE NcoI - EcoRI FRAGMENT
pMHS1 5' FLANK
ISOLATE EcoRI - SalI FRAGMENT FROM
pUC19 3'UTR/SV40
SUBCLONE IN pMH-1
CUT WITH NcoI - SalI
PstI
HindIII
EXON1
SmaI
BamHI
SalI
NcoI
-680
PS1 3' 5' FLANK
CAT
PstI FRAGMENT 4
BamHI (MADE BLUNT-ENDED
WITH KLENOW)
CUT WITH PstI
AND SmaI;
LIGATE WITH
FRAGMENT 4
PstI
RI
EXON1
CAT CODING
SEQUENCE
BamHI
3'UTR
HYBRID
αS1 CASEIN/
IgG SPLICE
NcoI
(-680)
SalI
pS13'5'CAT

FIG.-5B

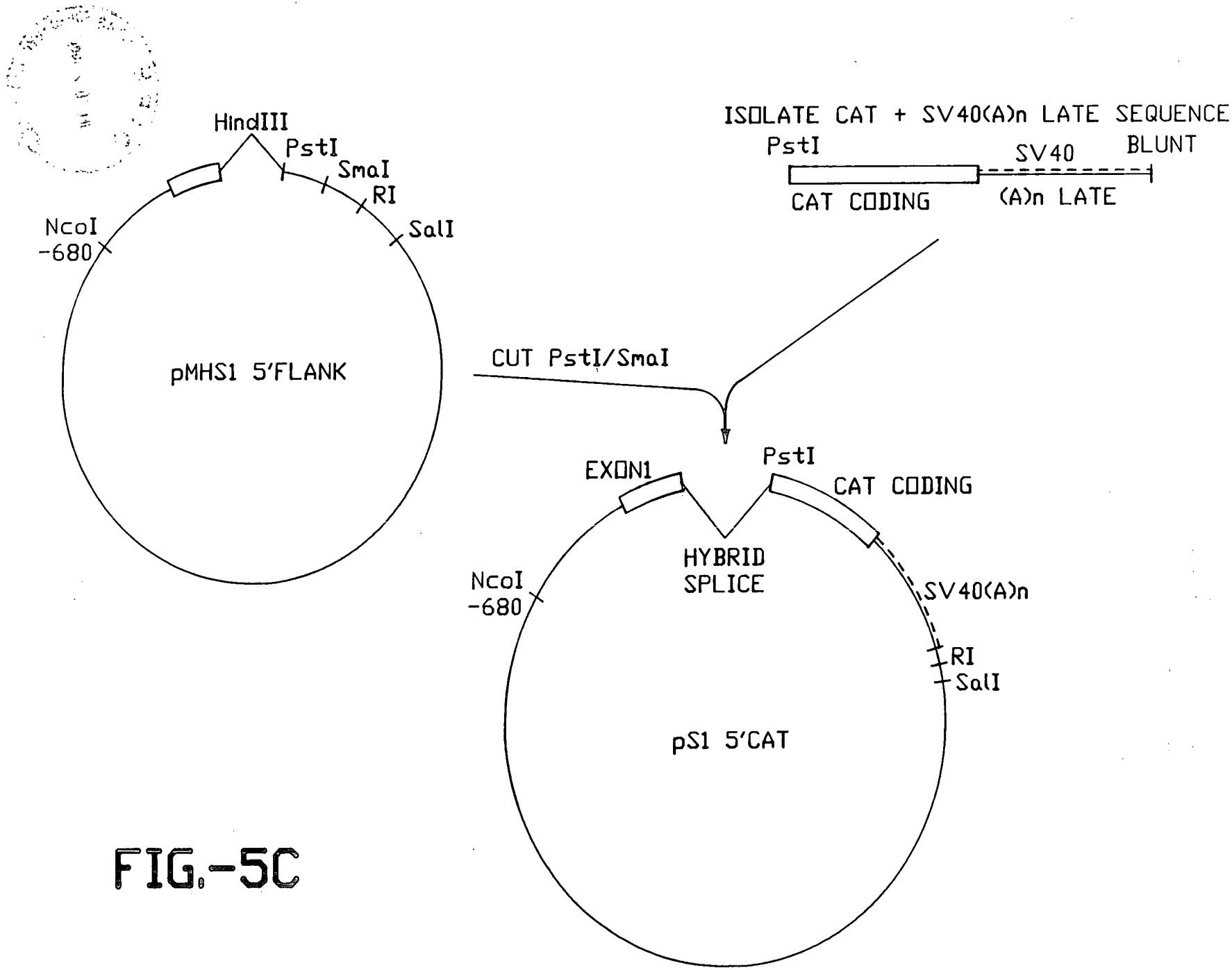

HindIII
PstI
SmaI
RI
SalI
NcoI
-680
pMHS1 5'FLANK
ISOLATE CAT + SV40(A)n LATE SEQUENCE
PstI
SV40
BLUNT
CAT CODING
(A)n LATE
CUT PstI/SmaI
EXON1
PstI
CAT CODING
HYBRID
SPLICE
NcoI
-680
SV40(A)n
RI
SalI
pS1 5'CAT

FIG.-5C

NAME pMH-1

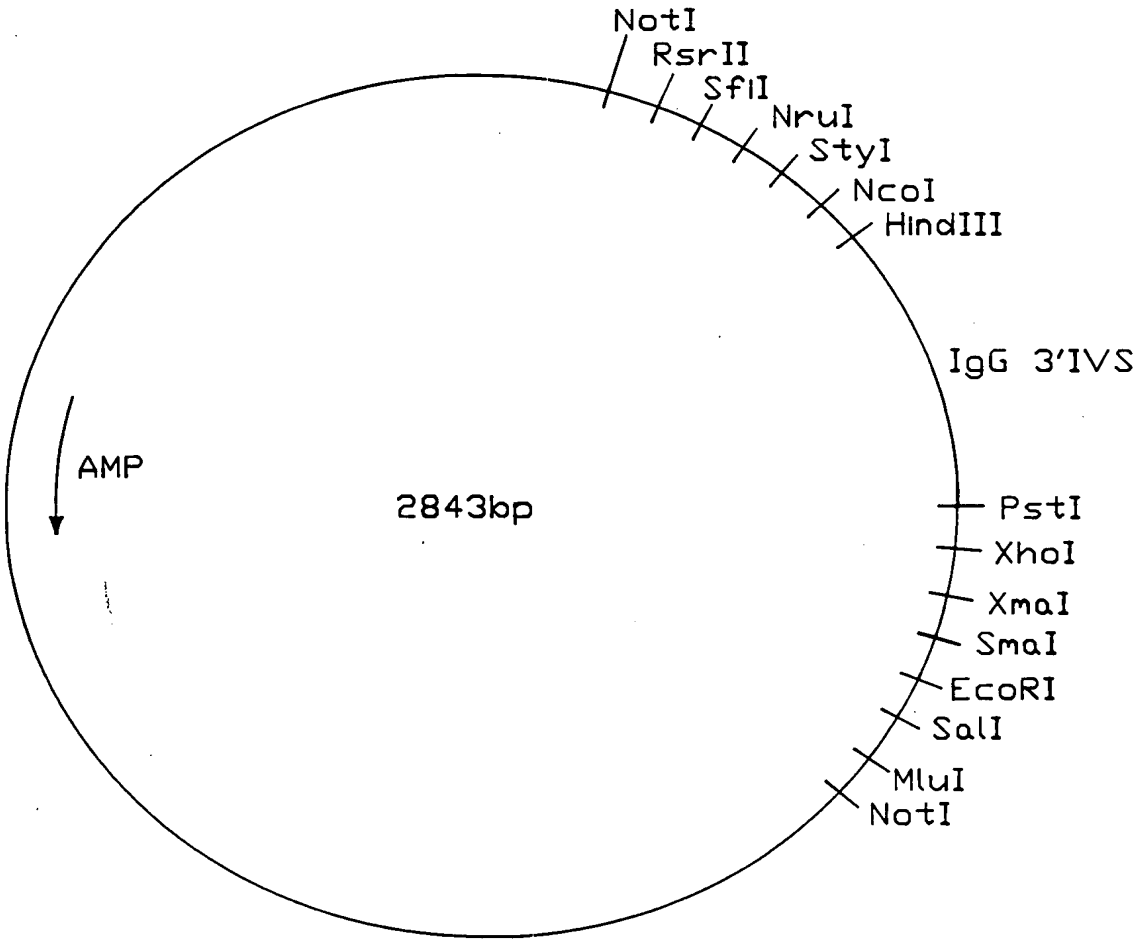
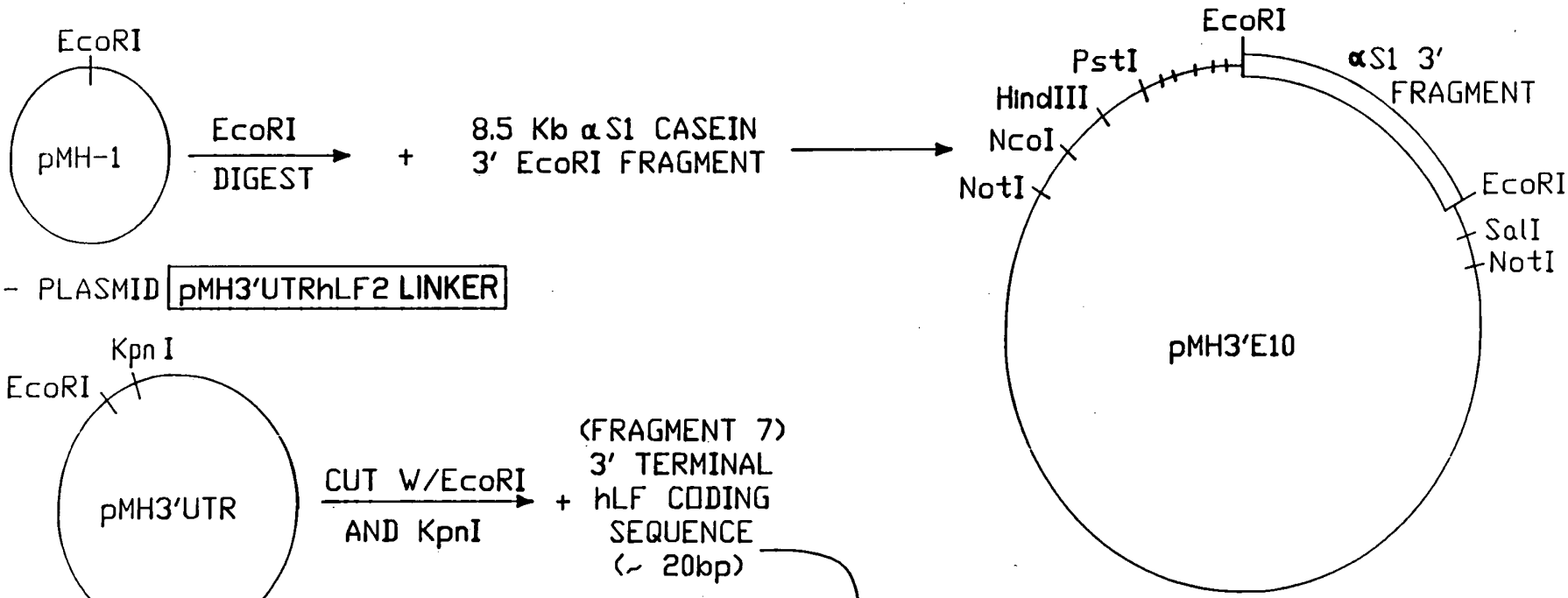


FIG.-6

- PLASMID pMH3'E10 CONTAINS ~ 8.5Kb EcoRI FRAGMENT OF THE 3'-END OF αS1 CASEIN. pMH-1 WAS CUT W/EcoRI + LIGATED TO THIS FRAGMENT



- PLASMID pMH3'UTRhLF2 LINKER

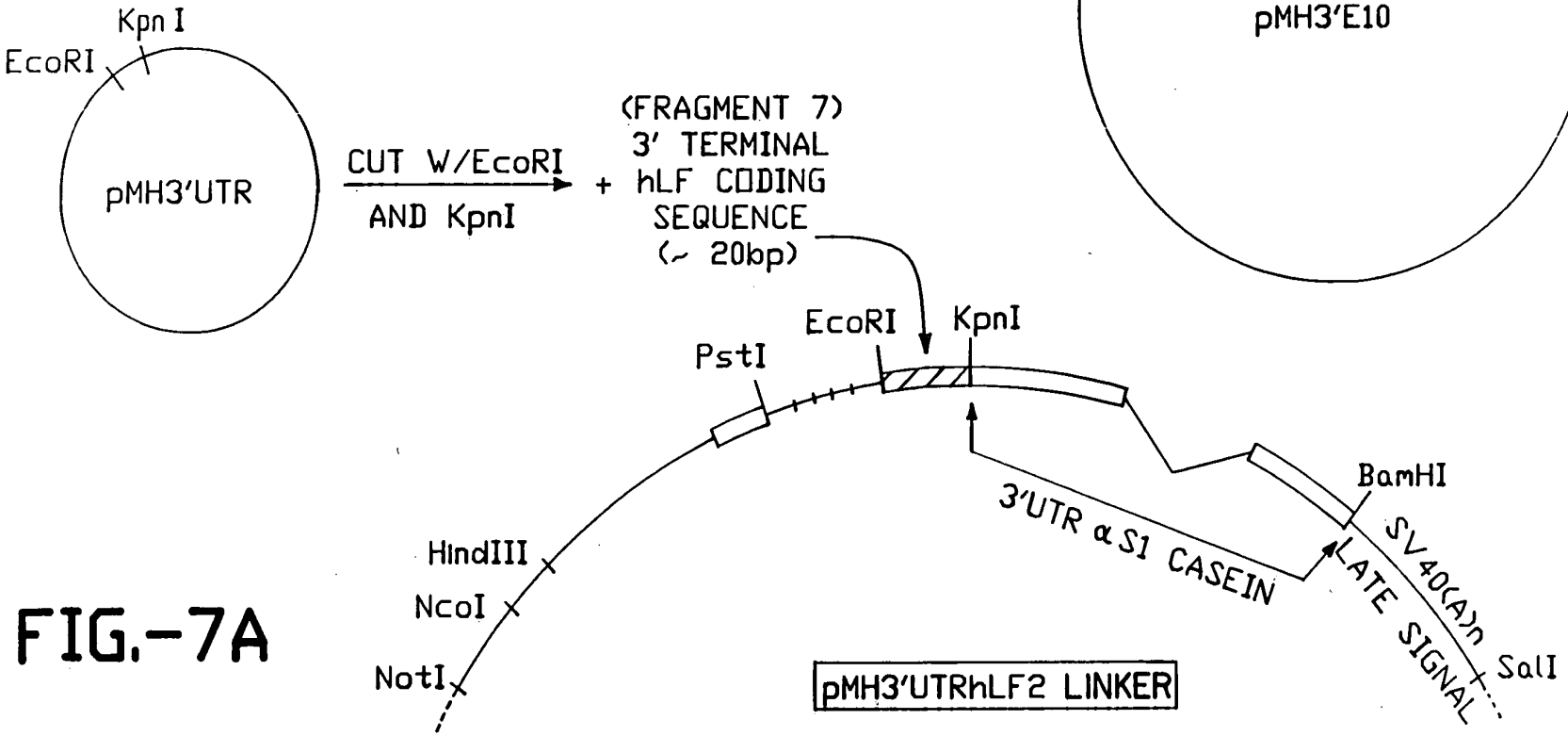
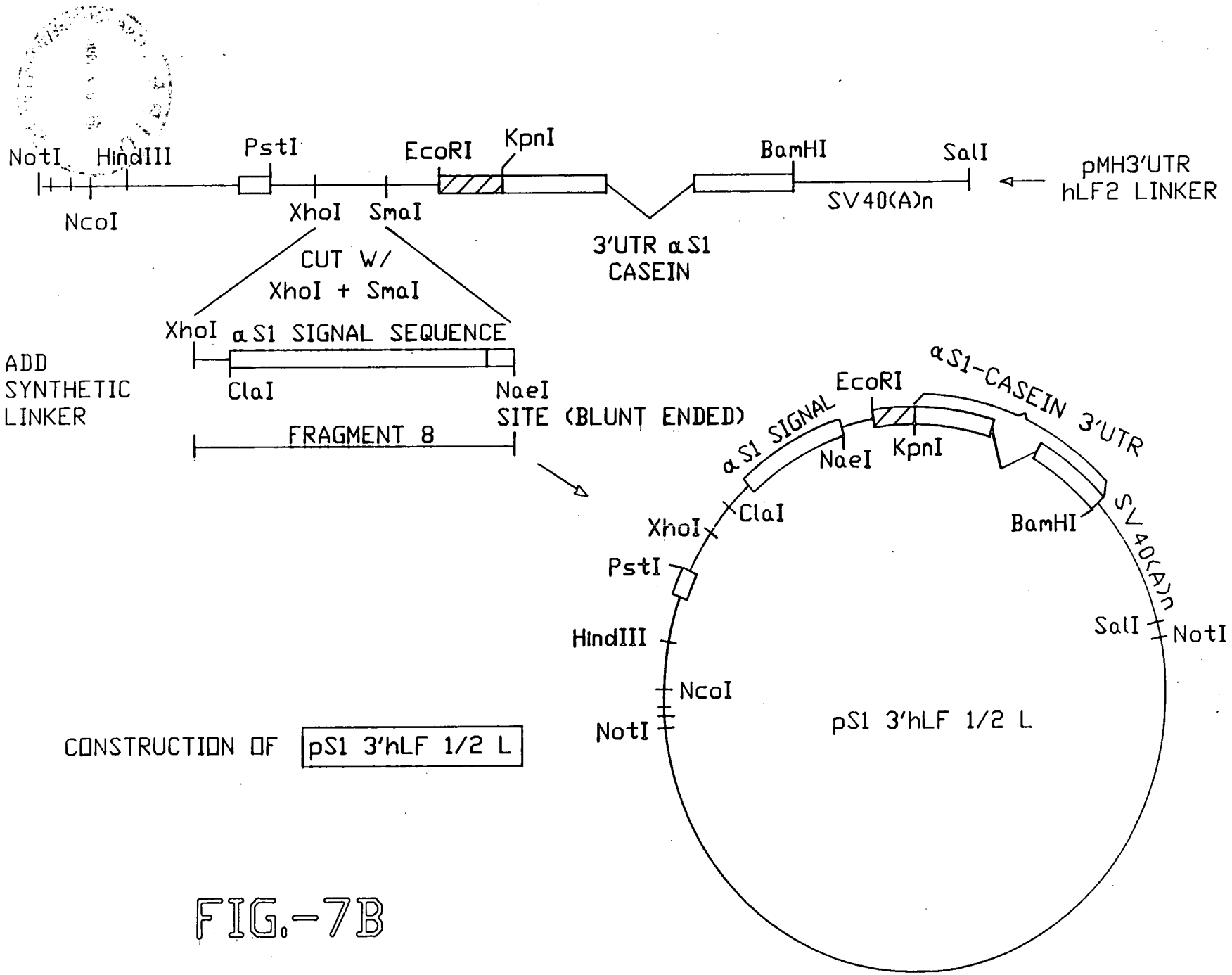


FIG.-7A



CONSTRUCTION OF pS1 3'hLF 1/2 L

FIG.-7B

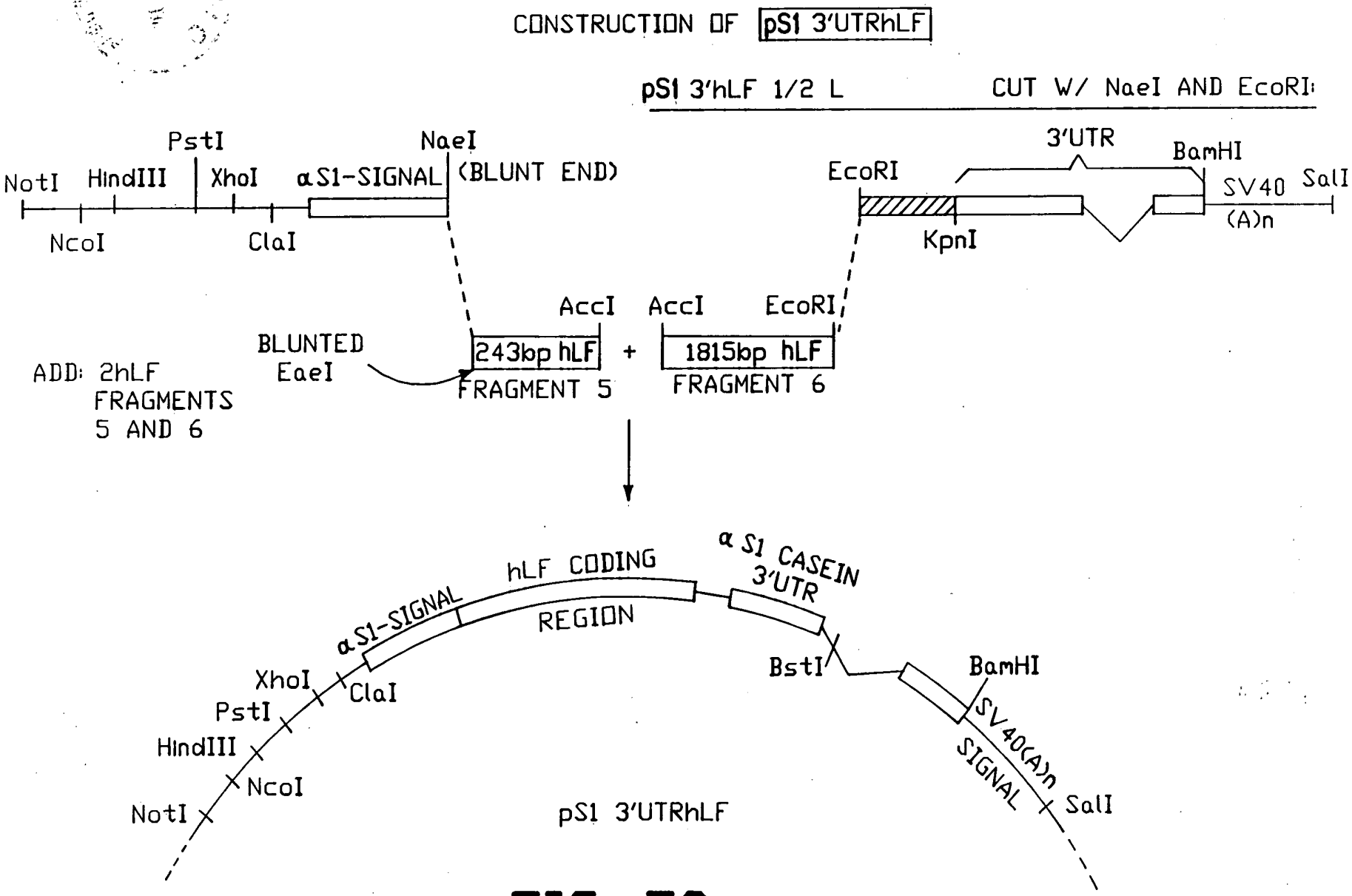

CONSTRUCTION OF pS1 3'UTRhLF
pS1 3'hLF 1/2 L
CUT W/ NaeI AND EcoRI:
PstI
NaeI
NotI
HindIII
XhoI
αS1-SIGNAL
(BLUNT END)
NcoI
ClaI
3'UTR
BamHI
EcoRI
SV40
SalI
(A)n
KpnI
AccI
AccI
EcoRI
BLUNTED
EaeI
243bp hLF
+
1815bp hLF
ADD: 2hLF
FRAGMENTS
5 AND 6
FRAGMENT 5
FRAGMENT 6
αS1 CASEIN
3'UTR
hLF CODING
REGION
αS1-SIGNAL
XhoI
ClaI
PstI
HindIII
NcoI
NotI
BstI
BamHI
SV40(A)n
SIGNAL
SalI
pS1 3'UTRhLF

FIG.-7C

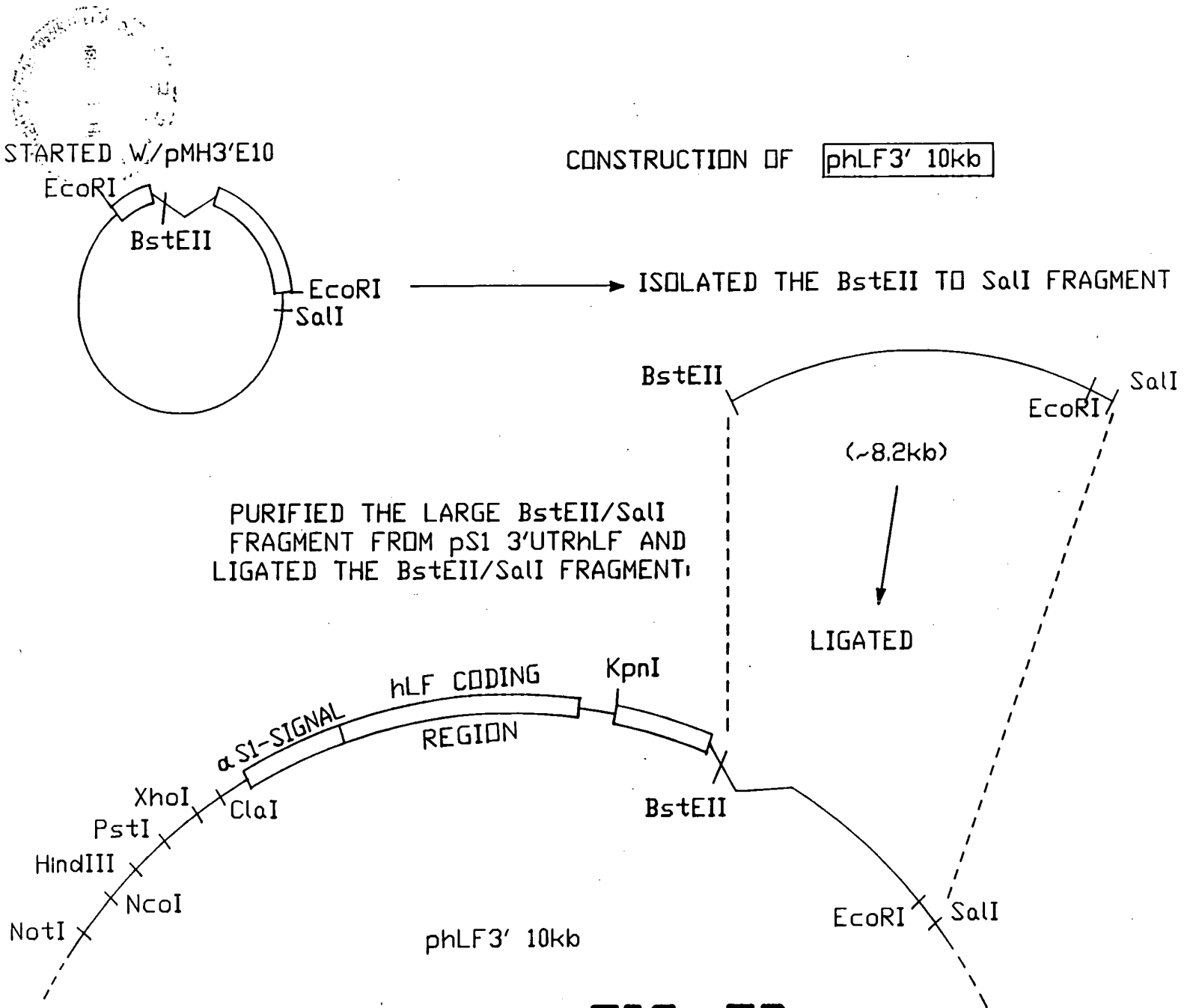
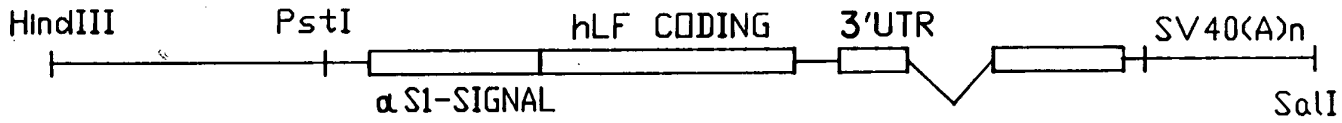

STARTED W/pMH3'E10
EcoRI
BstEII
EcoRI
SalI
CONSTRUCTION OF phLF3' 10kb
ISOLATED THE BstEII TO SalI FRAGMENT
BstEII
SalI
EcoRI
(~8.2kb)
PURIFIED THE LARGE BstEII/SalI FRAGMENT FROM pS1 3'UTRhLF AND LIGATED THE BstEII/SalI FRAGMENT
LIGATED
KpnI
hLF CODING REGION
αS1-SIGNAL
XhoI
ClaI
PstI
HindIII
NcoI
NotI
BstEII
EcoRI
SalI
phLF3' 10kb

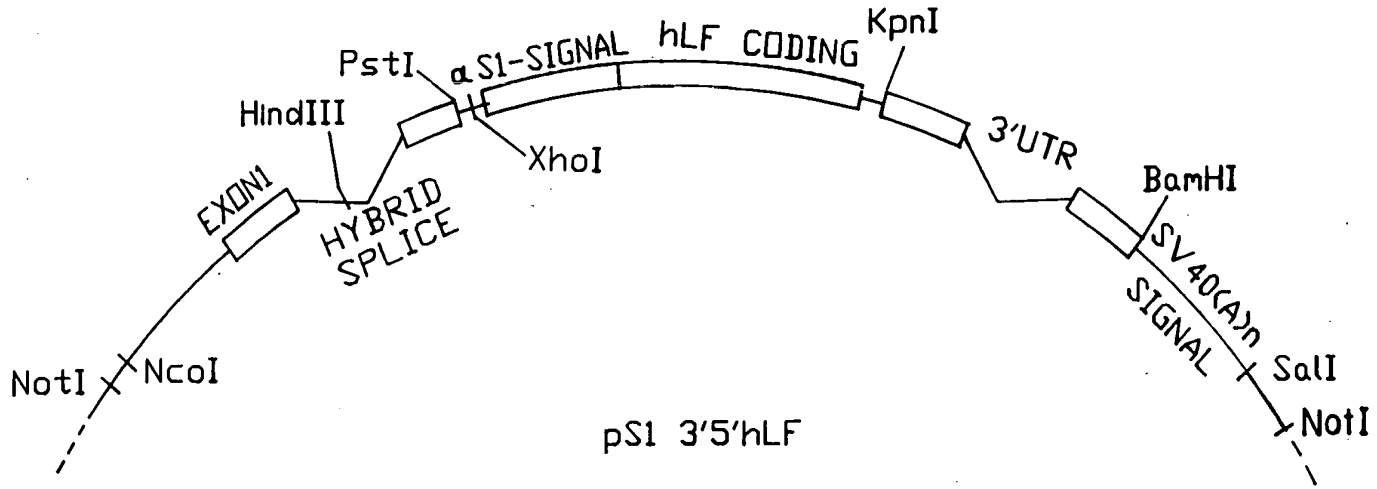
FIG.-7D

CONSTRUTION OF pS1 3'5'hLF

STARTED W/pS1 3'UTRhLF. CUT W/HindIII AND SalI AND PURIFIED THIS FRAGMENT CONTAINING THE αS1-CASEIN SIGNAL SEQUENCE, hLF CODING REGION, αS1UTR AND SV40(A)n.

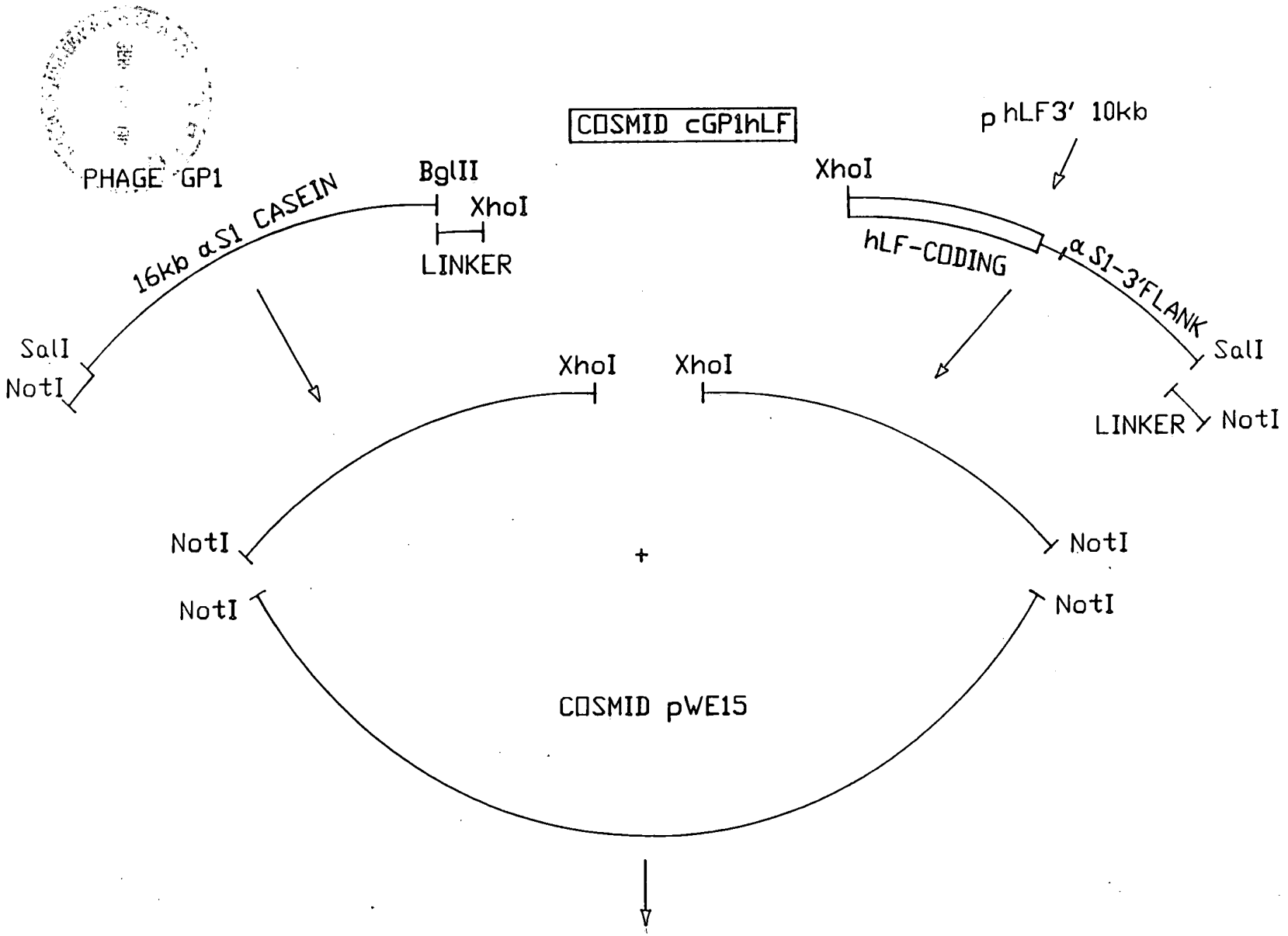


SUBCLONED INTO HindIII/SalI
CUT pMHS1 5' FLANK TO YIELD:



NOTE: pS1 5'hLF WAS MADE BY CUTTING pS13'5'hLF W/KpnI AND BamHI, FOLLOWED BY BLUNTING THE ENDS AND RELIGATING. THIS ELIMINATES THE SPLICED 3'UTR REGION.

FIG.-7E



3 WAY LIGATION. THE DNA FROM THIS COSMID IS PREPARED BY CUTTING WITH NotI AND PURIFYING THE EXPRESSION SEQUENCE PRIOR TO MICROINJECTION.

FIG.-7F

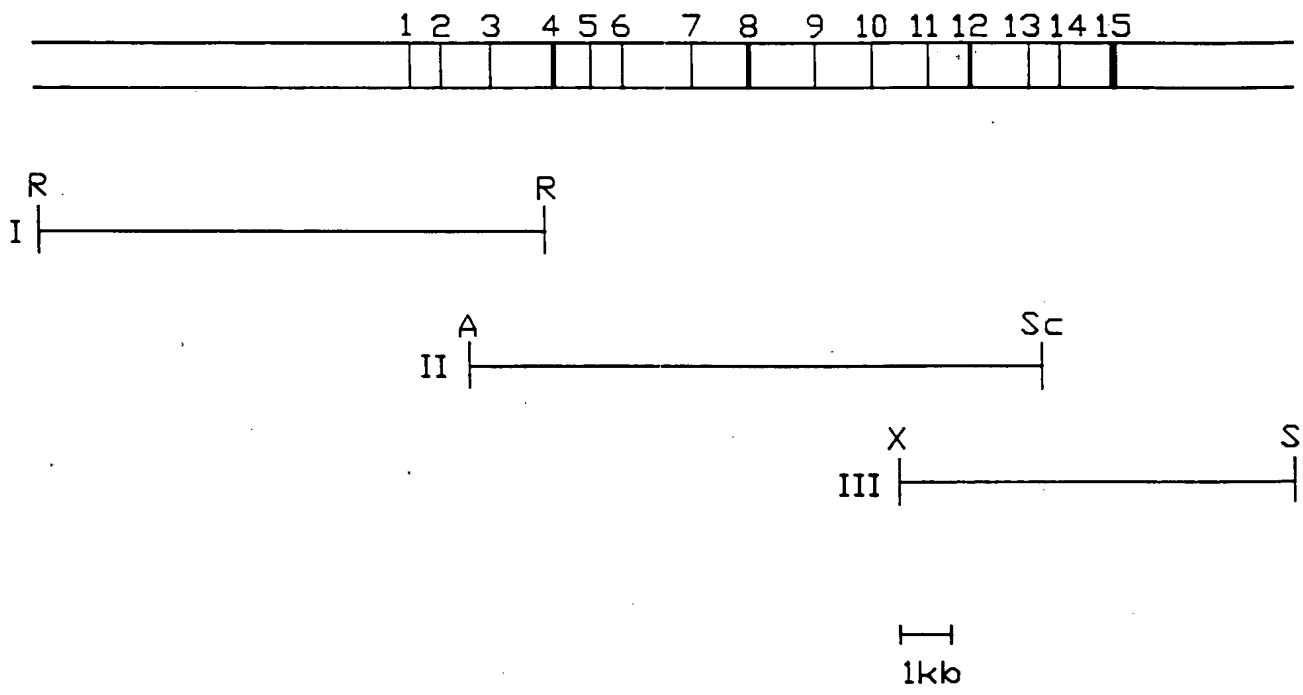

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
R
R
I
A
Sc
II
X
S
III
1kb

FIG.-8A

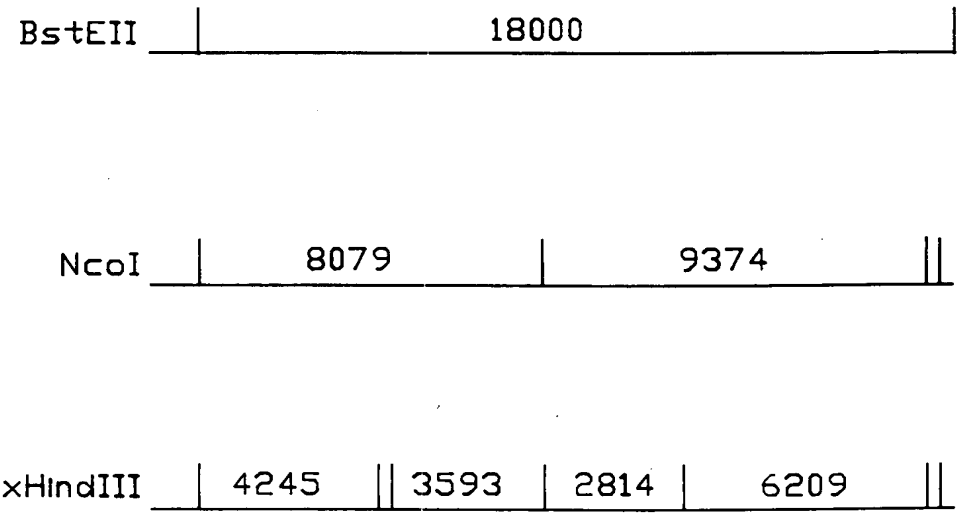

BstEII
18000
NcoI
8079
9374
NcoIxHindIII
4245
3593
2814
6209

FIG.-8B

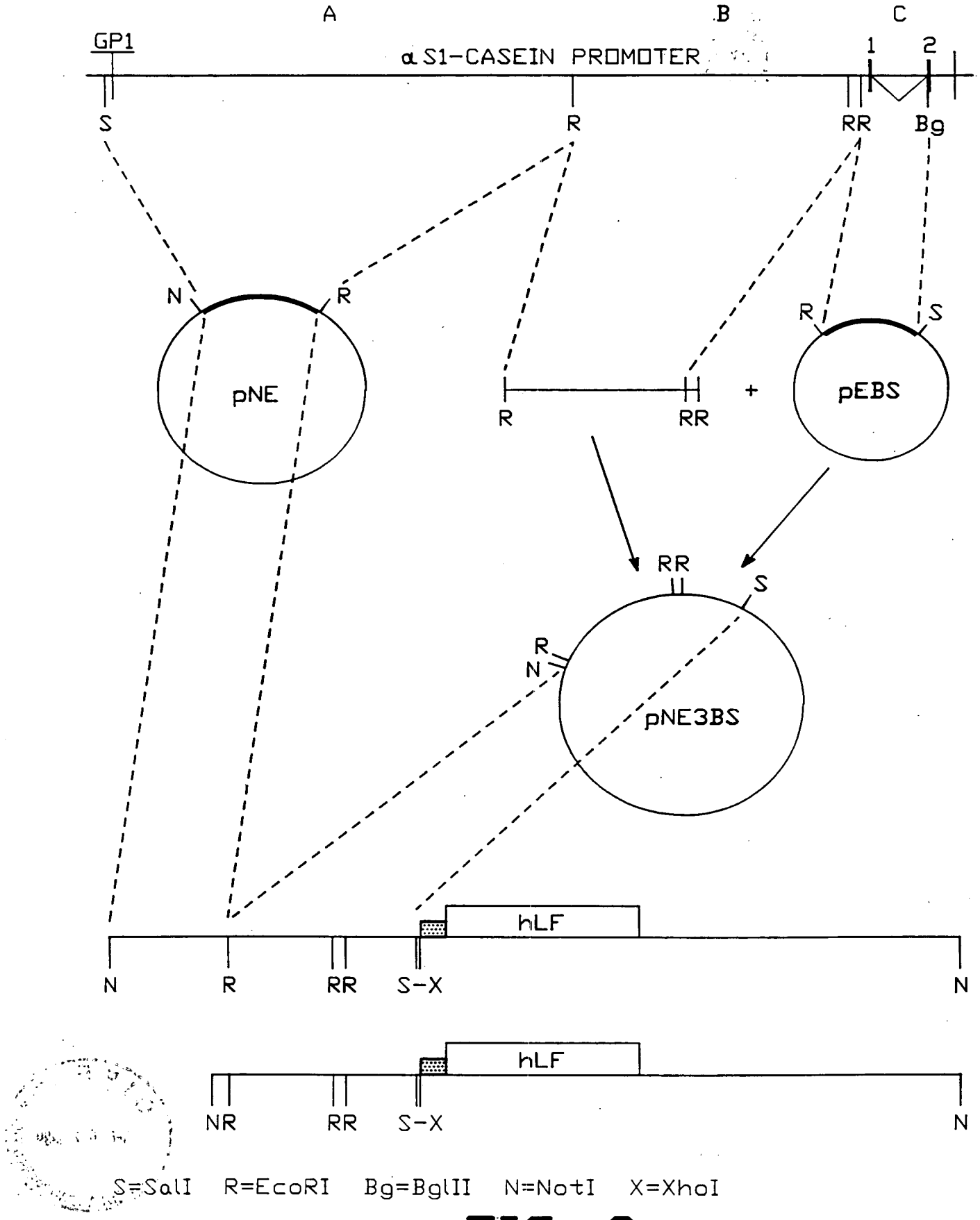

A
B
C
GP1
αS1-CASEIN PROMOTER
1
2
S
R
RR
Bg
N
R
pNE
R
RR
+
R
S
pEBS
RR
S
R
N
pNE3BS
hLF
N
R
RR
S-X
N
hLF
NR
RR
S-X
N
S=SalI R=EcoRI Bg=BglII N=NotI X=XhoI

FIG.-9

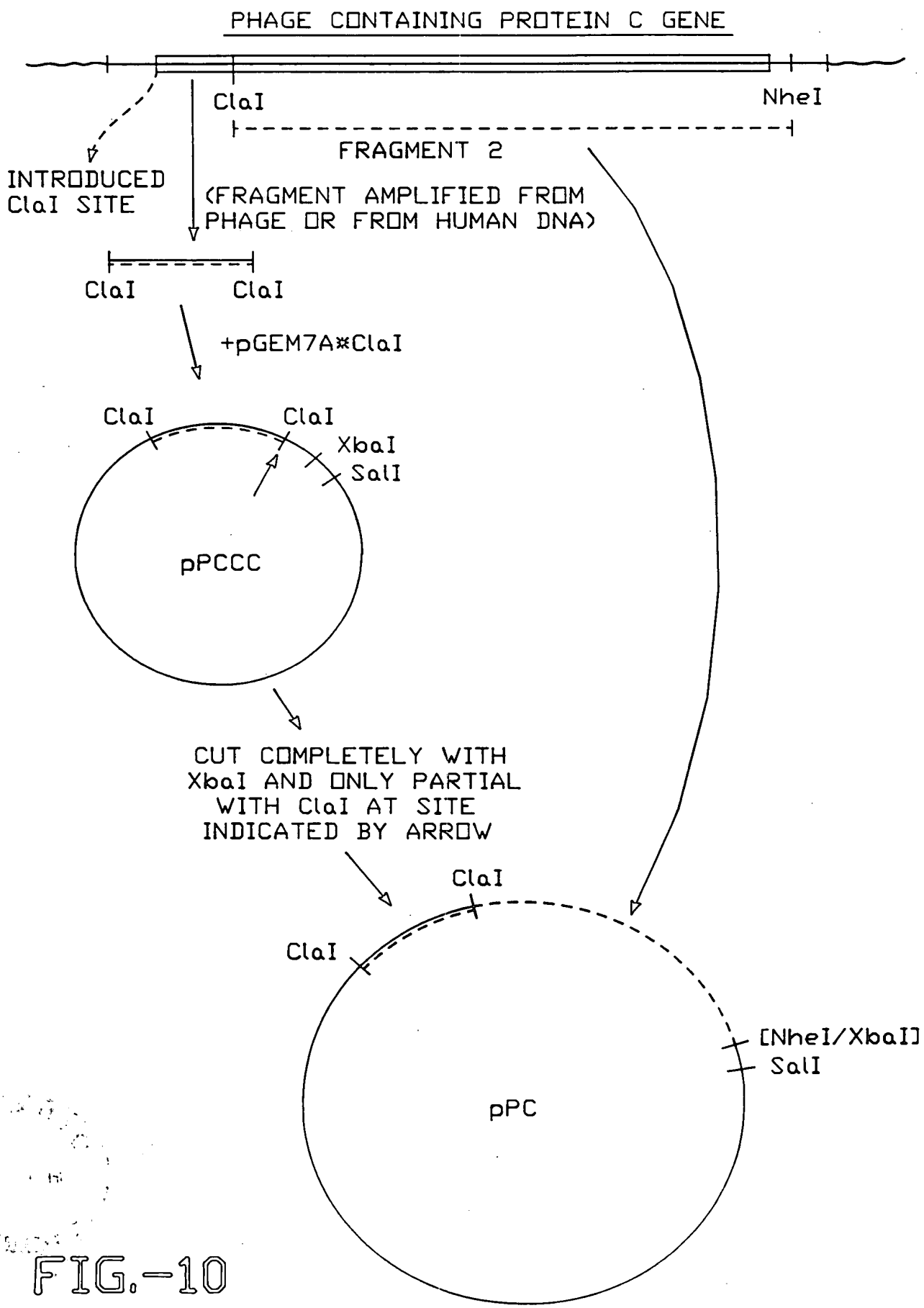

PHAGE CONTAINING PROTEIN C GENE
ClaI
NheI
FRAGMENT 2
INTRODUCED
ClaI SITE
(FRAGMENT AMPLIFIED FROM
PHAGE OR FROM HUMAN DNA)
ClaI
ClaI
+pGEM7A*ClaI
ClaI
ClaI
XbaI
SalI
pPCCC
CUT COMPLETELY WITH
XbaI AND ONLY PARTIAL
WITH ClaI AT SITE
INDICATED BY ARROW
ClaI
ClaI
[NheI/XbaI]
SalI
pPC

FIG.-10

5′- ATCACCTTGA TCATCAACCC AGCTTGCTGC TTCTTCCCAG

TCTTGGGTTC AAG gtattatgta tacatataac aaaatttcta tgattttcct ctgtctcatc

tttcattctt cactaatacg cagttgtaac ttttctatgt gattgcaagt attggtactt tcctatgata

tactgttagc aagcttgagg tgtggcaggc ttgagatctg gccatacact tgagtgacaa tgacatccac

tttgcctttc tctccacag GTGTCCACTC CCAGGTCCAA CTGCAG -3′

FIG.-11

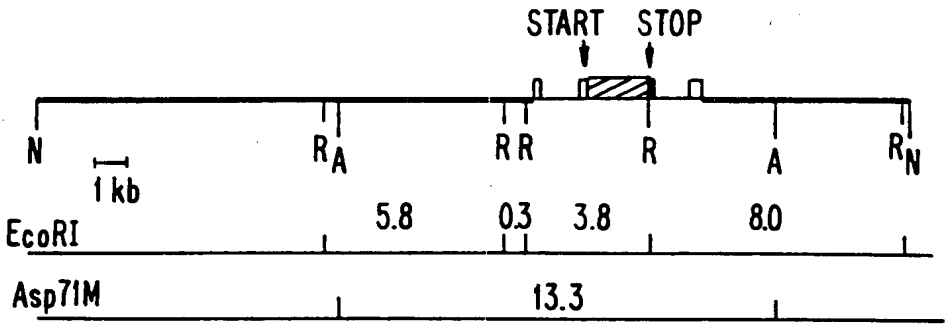


FIG. 12A.

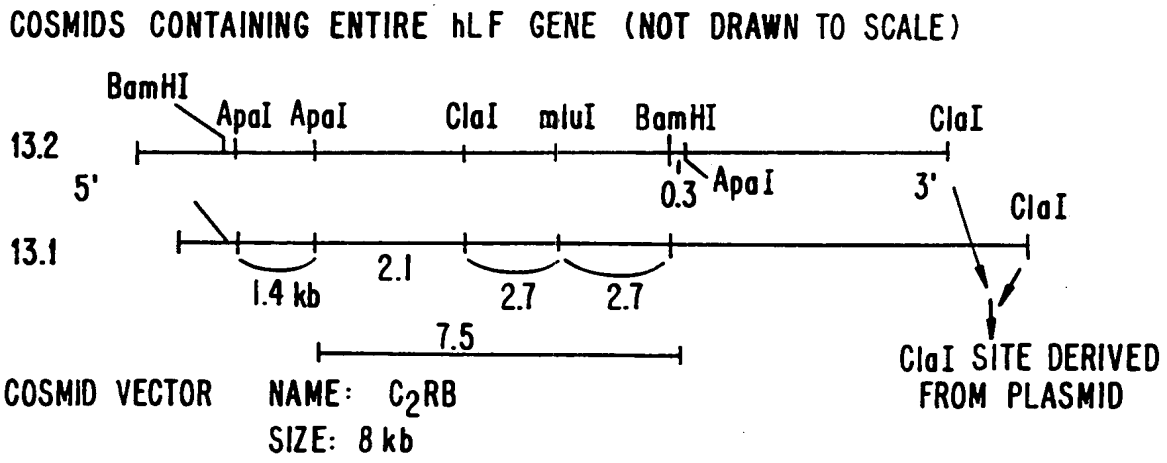


FIG. 13.

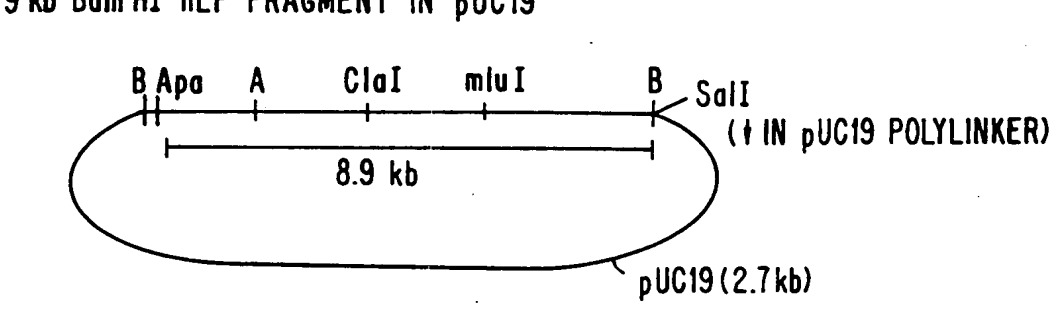


FIG. 14.

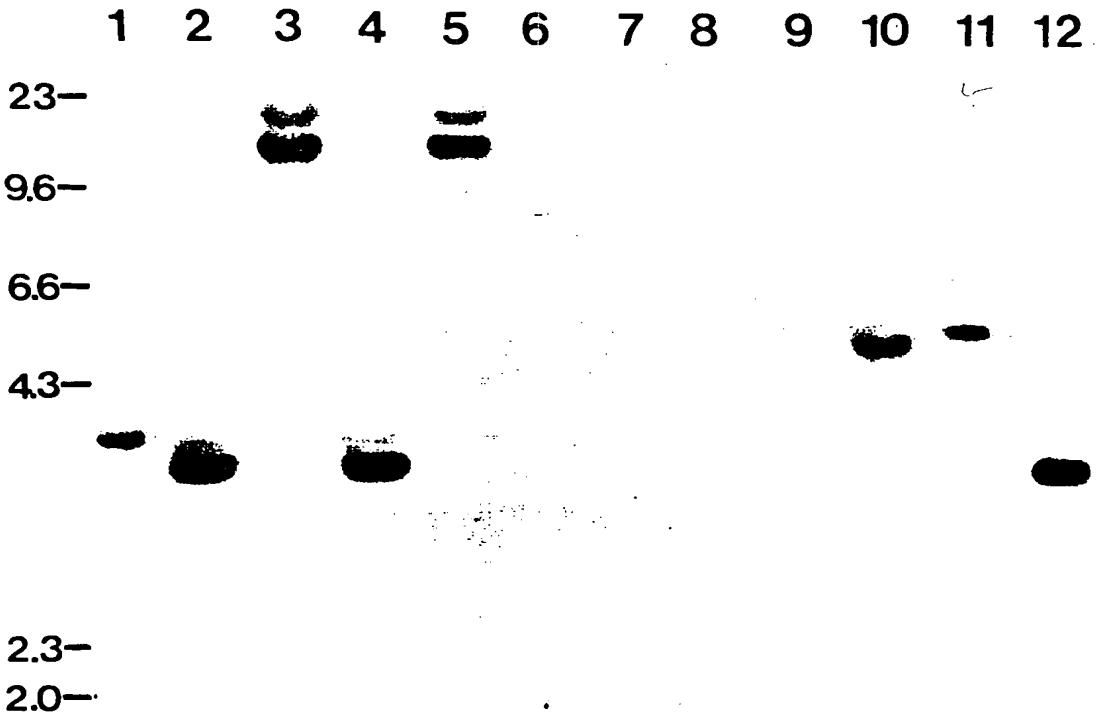
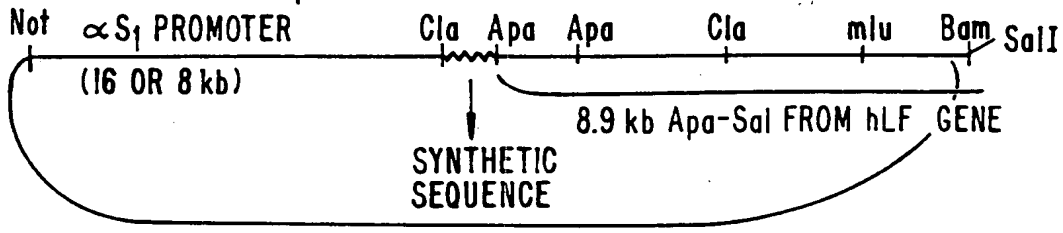


FIG. 12B.

LIGATION PRODUCT OF p16kbCS AND SYNTHETIC SEQ. (Cla-Apa)+ hLF FRAGMENT (p8kbCS)



CLONING VECTOR: pkUN (4 kb)

CONSTRUCT NAME: 8 hLF gen 9k, OR 16 hLF gen 9k

FIG. 15A.

STRUCTURE OF ClaI-ApaI SYNTHETIC SEQUENCE

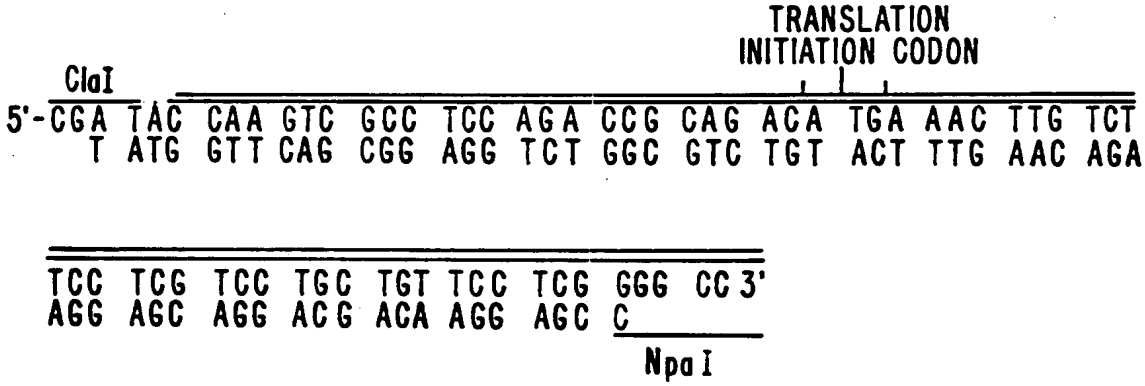
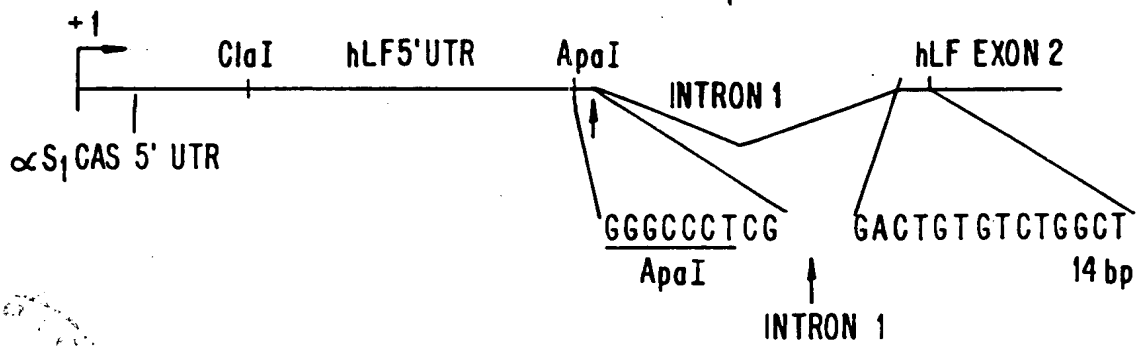


FIG. 15B.

+1: TRANSCRIPTION INITIATION SITE OF BOVINE αS1-CASEIN GENE



STRUCTURE OF REGION CONTAINING EXON 1 (HYBRID αS1-CASEIN/hLF EXON) AND PART OF EXON 2 OF THE GENOMIC hLF CONSTRUCTS DEPICTED IN FIGS. 15A THROUGH 17.

FIG. 15C.

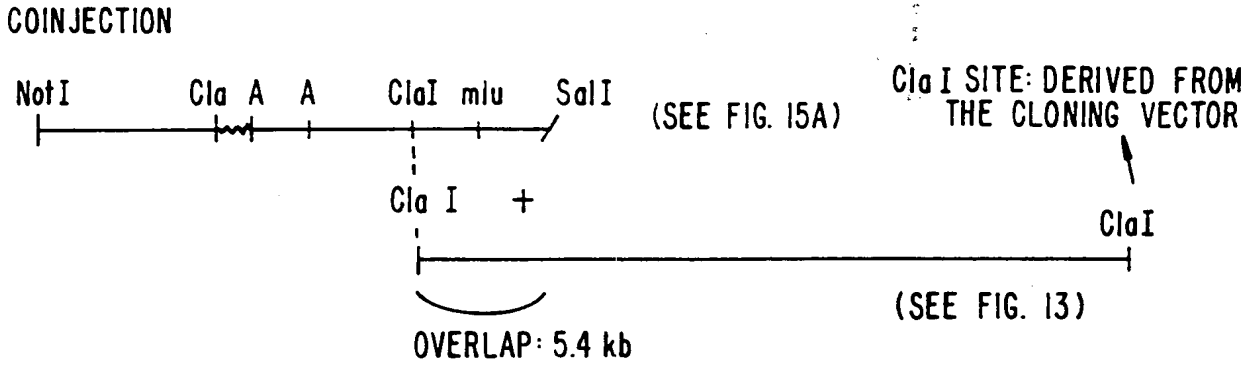

COINJECTION
NotI
Cla A A
ClaI mlu
SalI
(SEE FIG. 15A)
Cla I SITE: DERIVED FROM THE CLONING VECTOR
Cla I +
ClaI
(SEE FIG. 13)
OVERLAP: 5.4 kb

FIG. 16.

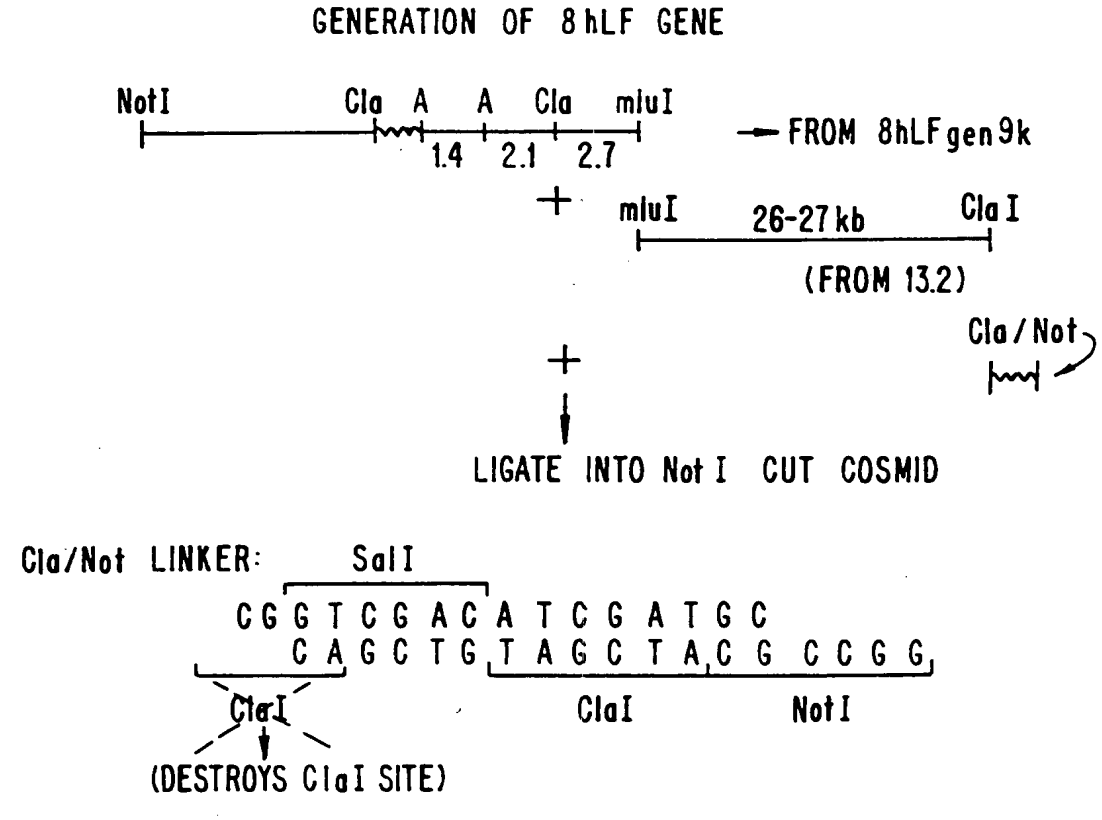

GENERATION OF 8 hLF GENE
NotI
Cla A A Cla mluI
1.4 2.1 2.7
FROM 8hLF gen 9k
+
mluI
26-27 kb
Cla I
(FROM 13.2)
Cla / Not
+
LIGATE INTO Not I CUT COSMID
Cla/Not LINKER:
SalI
C G G T C G A C A T C G A T G C
C A G C T G T A G C T A C G C C G G
ClaI
ClaI
NotI
(DESTROYS ClaI SITE)

FIG. 17.

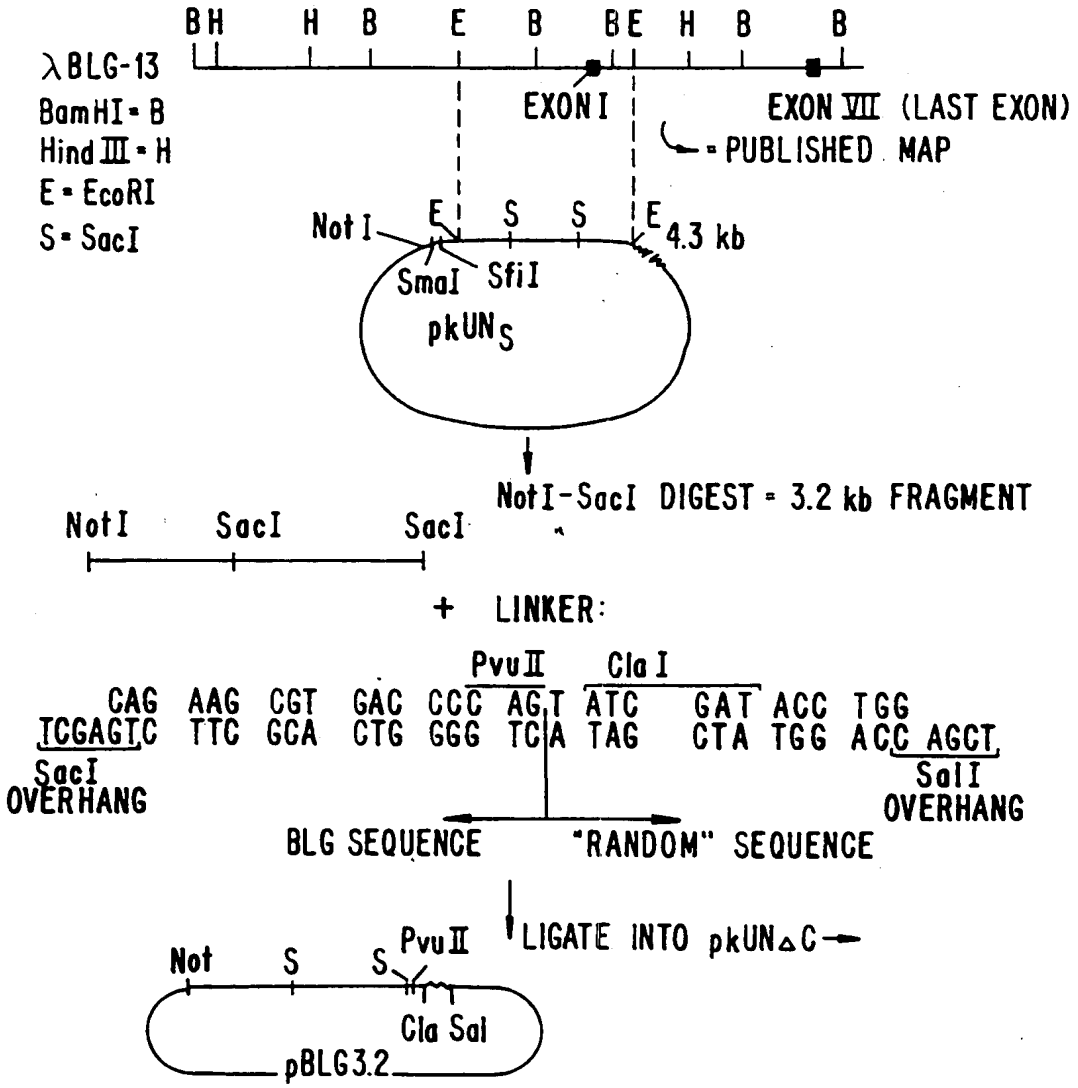

BH H B E B BE H B B
λBLG-13
BamHI = B
Hind III = H
E = EcoRI
S = SacI
EXON I
EXON VII (LAST EXON)
= PUBLISHED MAP
Not I
E S S E 4.3 kb
SmaI SfiI
pkUN S
NotI-SacI DIGEST = 3.2 kb FRAGMENT
NotI SacI SacI
+ LINKER:
PvuII ClaI
CAG AAG CGT GAC CCC AGT ATC GAT ACC TGG
TCGAGTC TTC GCA CTG GGG TCA TAG CTA TGG ACC AGCT
SacI
OVERHANG
SalI
OVERHANG
BLG SEQUENCE "RANDOM" SEQUENCE
LIGATE INTO pkUNΔC
Not S S PvuII
Cla Sal
pBLG3.2

FIG. 18.

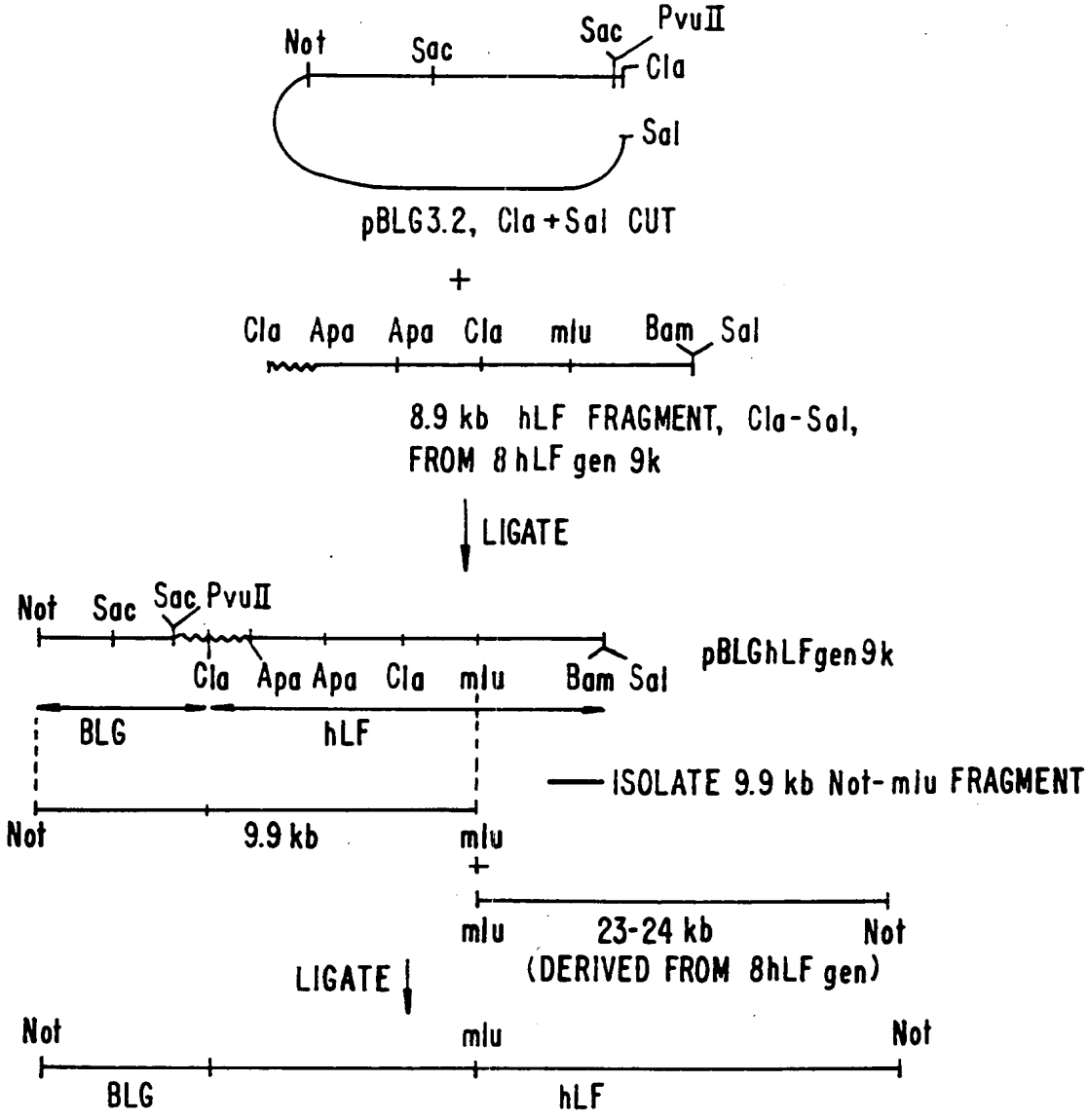

Not
Sac
Sac PvuII
Cla
Sal
pBLG3.2, Cla+Sal CUT
+
Cla Apa Apa Cla mlu Bam Sal
8.9 kb hLF FRAGMENT, Cla-Sal,
FROM 8 hLF gen 9k
LIGATE
Not Sac Sac PvuII
Cla Apa Apa Cla mlu Bam Sal
pBLGhLFgen9k
BLG
hLF
ISOLATE 9.9 kb Not-mlu FRAGMENT
Not
9.9 kb
mlu
+
mlu
23-24 kb
Not
(DERIVED FROM 8hLF gen)
LIGATE
Not
mlu
Not
BLG
hLF

FIG. 19.

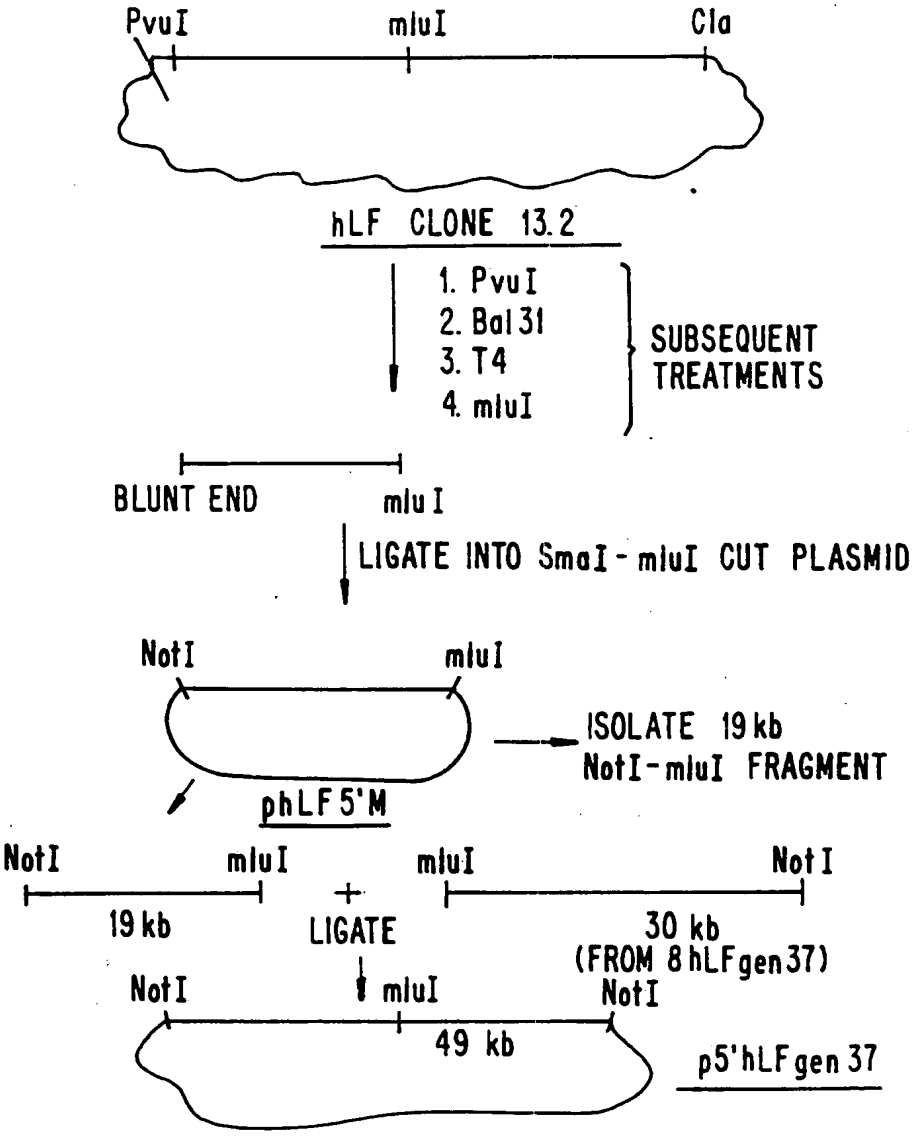
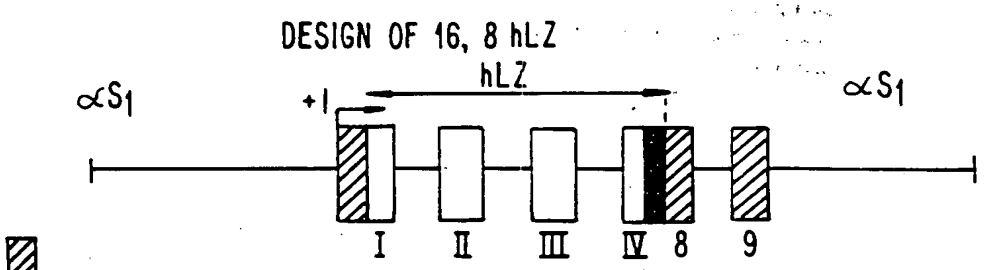

PvuI
mluI
Cla
hLF CLONE 13.2
1. PvuI
2. Bal 31
3. T4
4. mluI
SUBSEQUENT TREATMENTS
BLUNT END
mlu I
LIGATE INTO SmaI - mluI CUT PLASMID
NotI
mluI
ISOLATE 19 kb NotI - mluI FRAGMENT
phLF 5'M
NotI
mluI
mluI
NotI
19 kb
LIGATE
30 kb (FROM 8 hLFgen37)
NotI
mluI
NotI
49 kb
p5'hLFgen 37

FIG. 20.



: EXON FROM αS_1 - CASEIN GENE (NON-CODING)

: EXON FROM hLZ GENE

+1: TRANSCRIPTION INITIATION SITE

: NON-CODING PART OF hLZ EXON 4

I, 8 AND 9: αS_1 - GENE EXON NUMBERS

I-IV: NUMBERS OF hLZ EXONS

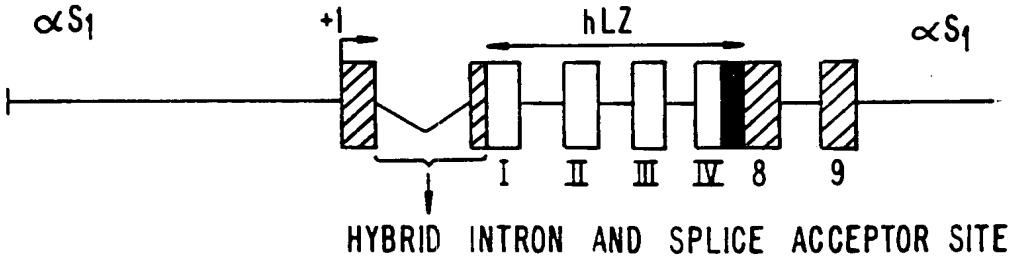
FLANKING REGIONS: FROM αS_1 CASEIN GENE

(SIGNAL SEQUENCE: hLZ)

FIG. 21.

DESIGN OF 16, 8 hLZ3

LEGEND: SEE FIGURE 9



(SIGNAL SEQUENCE: hLZ)

FIG. 22.

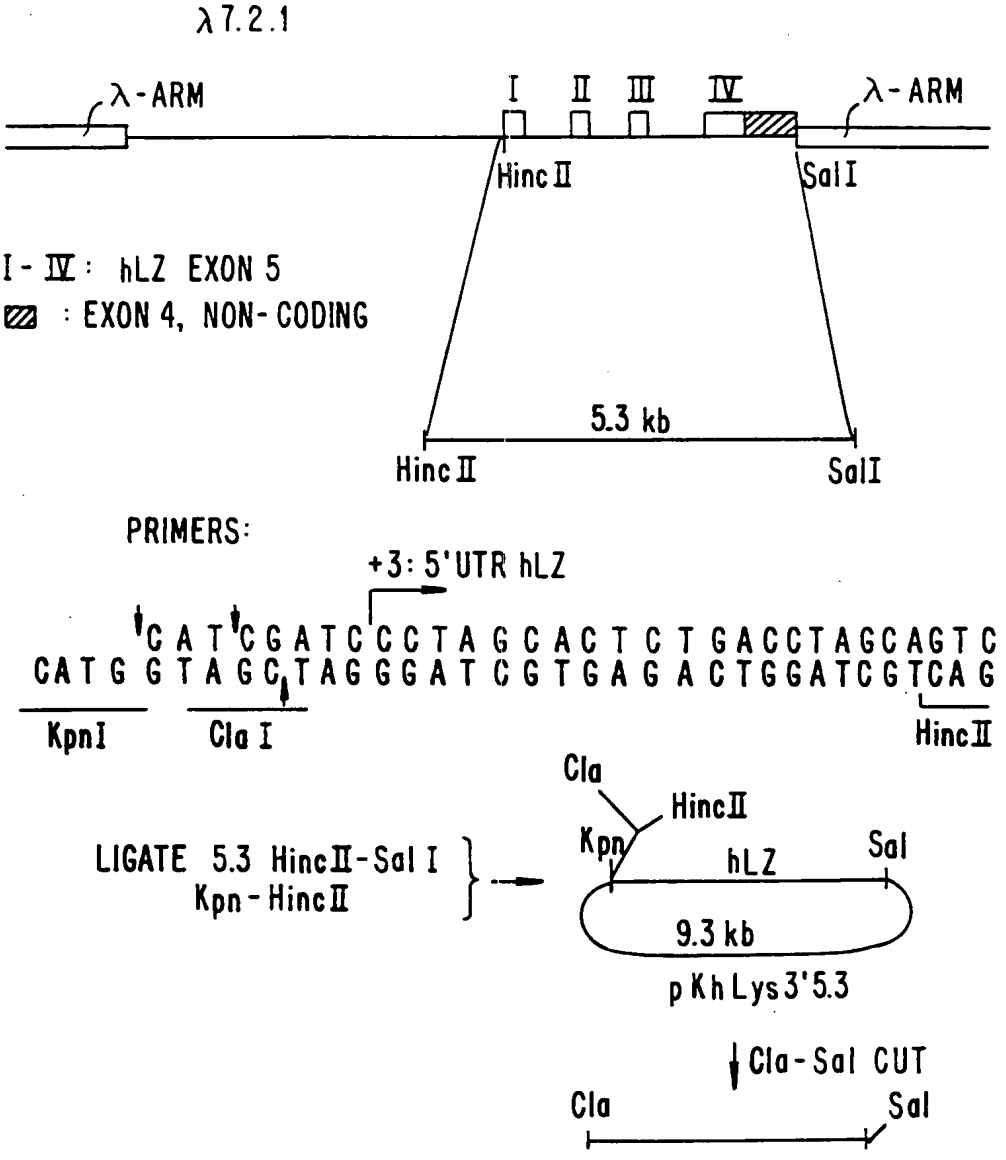

λ7.2.1
λ-ARM
I II III IV
λ-ARM
Hinc II
Sal I
I - IV: hLZ EXON 5
: EXON 4, NON-CODING
5.3 kb
Hinc II
Sal I
PRIMERS:
+3: 5'UTR hLZ
C A T C G A T C C C T A G C A C T C T G A C C T A G C A G T C
C A T G G T A G C T A G G G A T C G T G A G A C T G G A T C G T C A G
Kpn I
Cla I
Hinc II
Cla
Hinc II
Kpn
hLZ
Sal
LIGATE 5.3 Hinc II-Sal I
Kpn-Hinc II
9.3 kb
pKhLys3'5.3
Cla-Sal CUT
Cla
Sal

FIG. 23A.

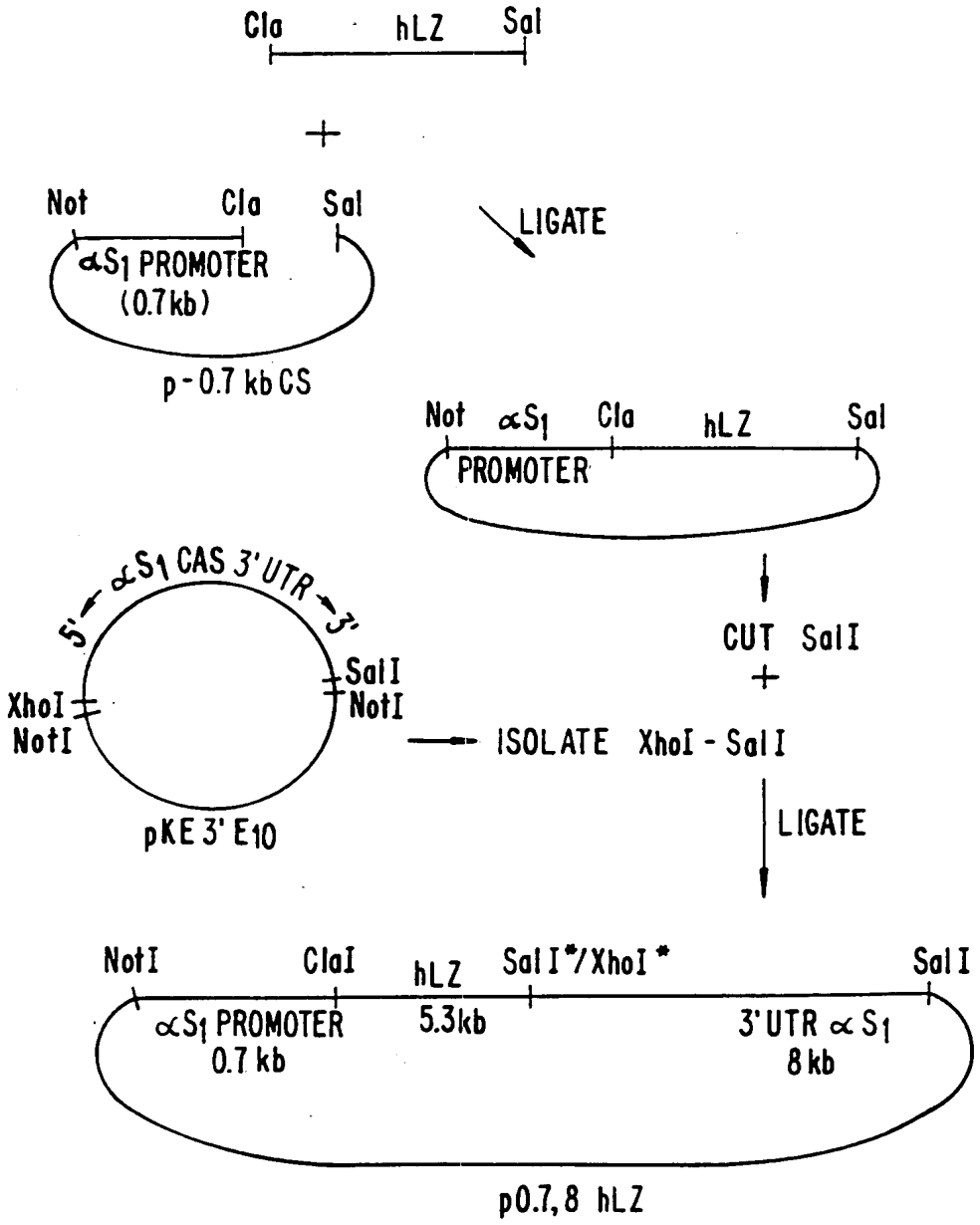

Cla
hLZ
Sal
+
Not
Cla
Sal
LIGATE
αS1 PROMOTER
(0.7kb)
p-0.7 kb CS
Not
αS1
Cla
hLZ
Sal
PROMOTER
5'
αS1 CAS 3'UTR
3'
SalI
NotI
XhoI
NotI
CUT SalI
+
ISOLATE XhoI - SalI
LIGATE
pKE 3' E10
NotI
ClaI
hLZ
SalI*/XhoI*
SalI
αS1 PROMOTER
0.7 kb
5.3kb
3'UTR αS1
8 kb
p0.7,8 hLZ

FIG. 23B.

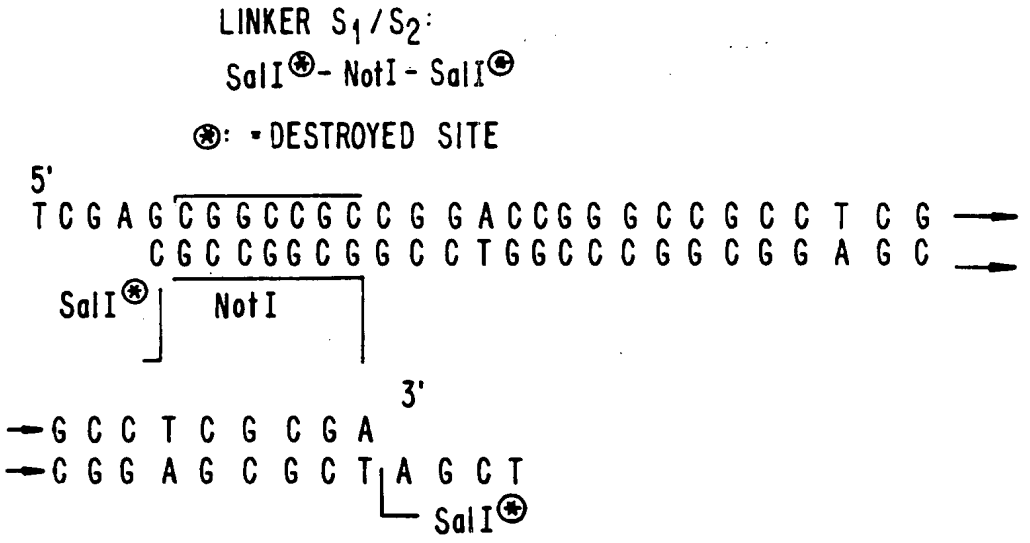


FIG. 23C.

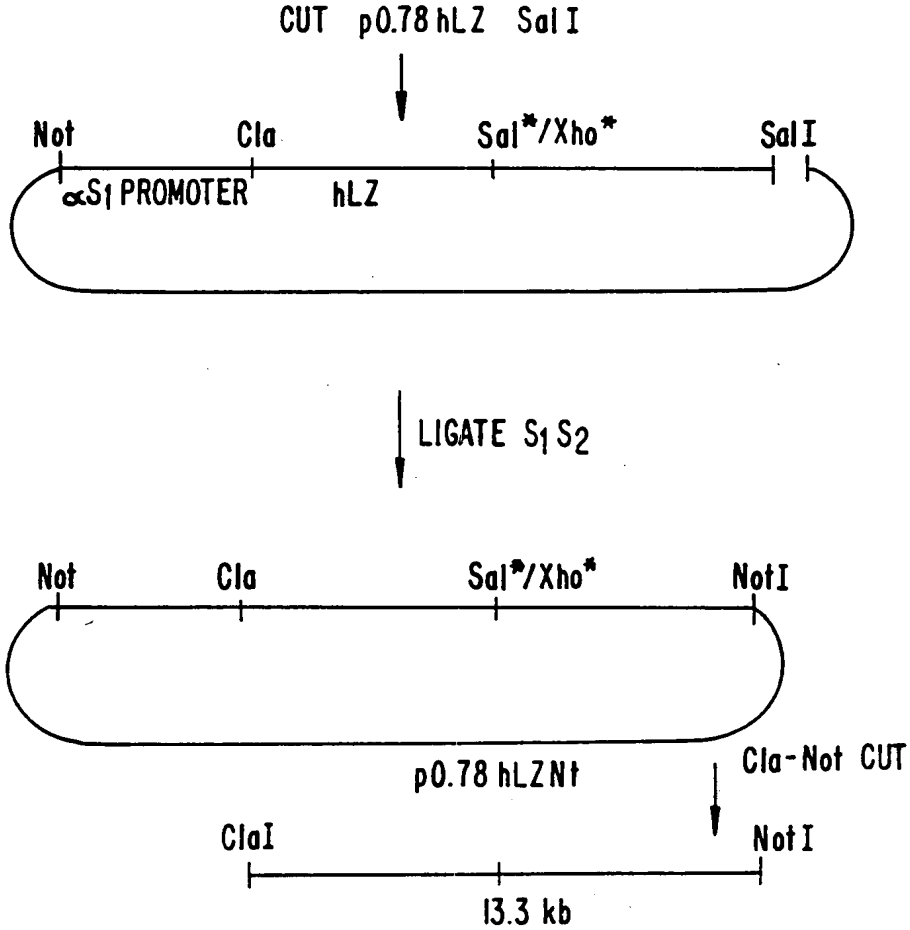


FIG. 23D.

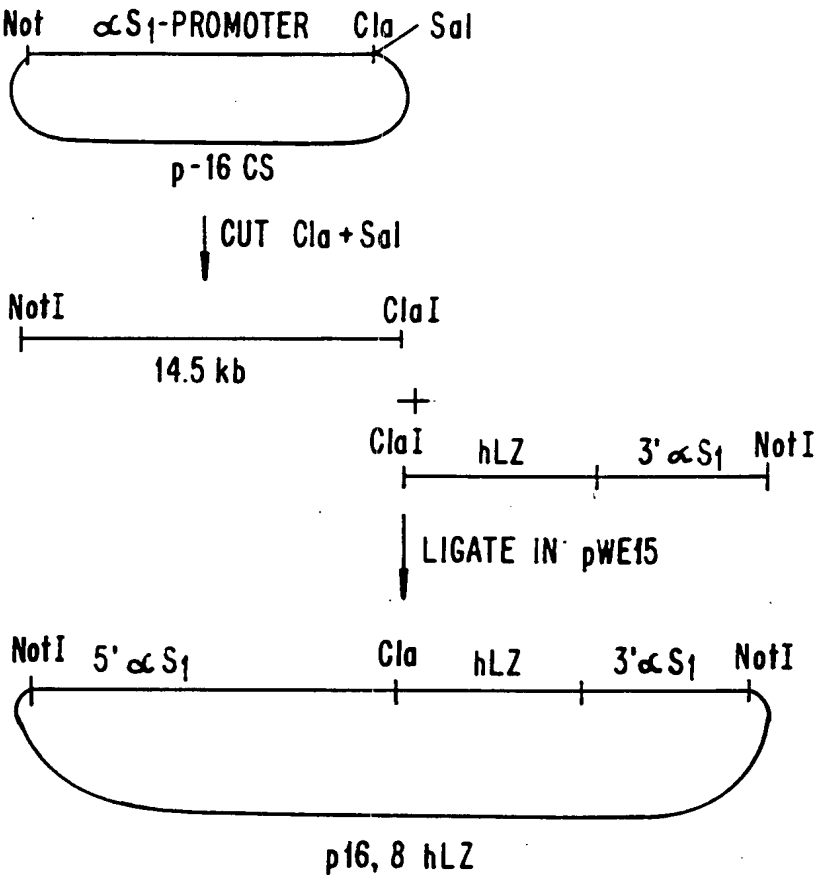

Not
αS1-PROMOTER
Cla
Sal
p-16 CS
CUT Cla + Sal
NotI
ClaI
14.5 kb
+
ClaI
hLZ
3'αS1
NotI
LIGATE IN pWE15
NotI
5'αS1
Cla
hLZ
3'αS1
NotI
p16, 8 hLZ

FIG. 23E.

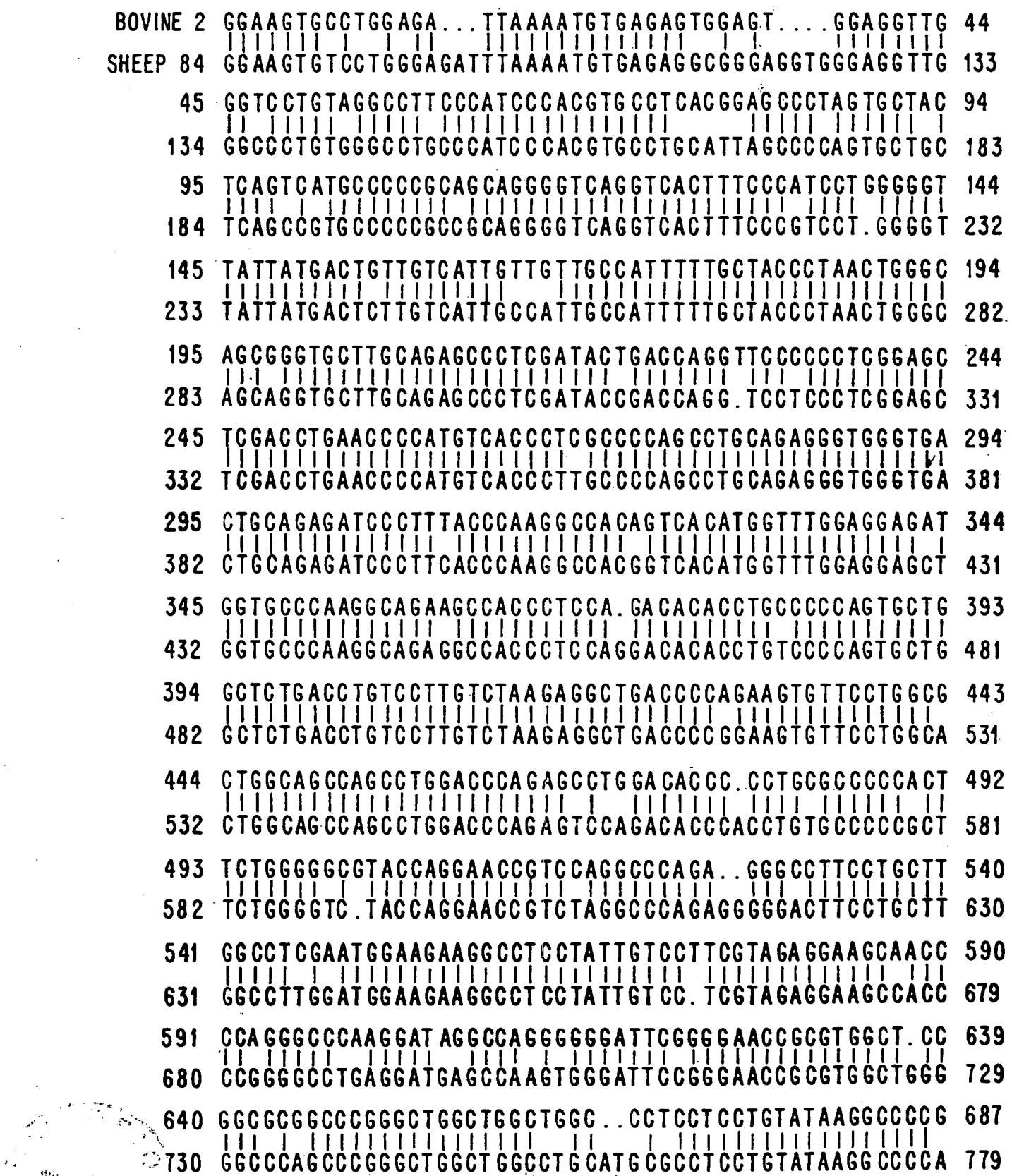

BOVINE 2 GGAAGTGCCTGGAGA...TTAAAATGTGAGAGTGGAGT....GGAGGTTG 44
SHEEP 84 GGAAGTGTCCTGGGAGATTTAAAATGTGAGAGGCGGGAGGTGGGAGGTTG 133
45 GGTCCTGTAGGCCTTCCCATCCCACGTGCCTCACGGAGCCCTAGTGCTAC 94
134 GGCCCTGTGGGCCTGCCCATCCCACGTGCCTGCATTAGCCCCAGTGCTGC 183
95 TCAGTCATGCCCCCGCAGCAGGGGTCAGGTCACTTTCCCATCCTGGGGGT 144
184 TCAGCCGTGCCCCCGCCGCAGGGGTCAGGTCACTTTCCCGTCCT.GGGGT 232
145 TATTATGACTGTTGTCATTGTTGTTGCCATTTTTGCTACCCTAACTGGGC 194
233 TATTATGACTCTTGTCATTGCCATTGCCATTTTTGCTACCCTAACTGGGC 282
195 AGCGGGTGCTTGCAGAGCCCTCGATACTGACCAGGTTCCCCCCTCGGAGC 244
283 AGCAGGTGCTTGCAGAGCCCTCGATACCGACCAGG.TCCTCCCTCGGAGC 331
245 TCGACCTGAACCCCATGTCACCCTCGCCCCAGCCTGCAGAGGGTGGGTGA 294
332 TCGACCTGAACCCCATGTCACCCTTGCCCCAGCCTGCAGAGGGTGGGTGA 381
295 CTGCAGAGATCCCTTTACCCAAGGCCACAGTCACATGGTTTGGAGGAGAT 344
382 CTGCAGAGATCCCTTCACCCAAGGCCACGGTCACATGGTTTGGAGGAGCT 431
345 GGTGCCCAAGGCAGAAGCCACCCTCCA.GACACACCTGCCCCCAGTGCTG 393
432 GGTGCCCAAGGCAGAGGCCACCCTCCAGGACACACCTGTCCCCAGTGCTG 481
394 GCTCTGACCTGTCCTTGTCTAAGAGGCTGACCCCAGAAGTGTTCCTGGCG 443
482 GCTCTGACCTGTCCTTGTCTAAGAGGCTGACCCCGGAAGTGTTCCTGGCA 531
444 CTGGCAGCCAGCCTGGACCCAGAGCCTGGACACCC.CCTGCGCCCCCACT 492
532 CTGGCAGCCAGCCTGGACCCAGAGTCCAGACACCCACCTGTGCCCCCGCT 581
493 TCTGGGGGCGTACCAGGAACCGTCCAGGCCCAGA..GGGCCTTCCTGCTT 540
582 TCTGGGGTC.TACCAGGAACCGTCTAGGCCCAGAGGGGGACTTCCTGCTT 630
541 GGCCTCGAATGGAAGAAGGCCTCCTATTGTCCTTCGTAGAGGAAGCAACC 590
631 GGCCTTGGATGGAAGAAGGCCTCCTATTGTCC.TCGTAGAGGAAGCCACC 679
591 CCAGGGCCCAAGGATAGGCCAGGGGGGATTCGGGGAACCGCGTGGCT.CC 639
680 CCGGGGCCTGAGGATGAGCCAAGTGGGATTCCGGGAACCGCGTGGCTGGG 729
640 GGCGCGGCCCGGGCTGGCTGGCTGGC..CCTCCTCCTGTATAAGGCCCCG 687
730 GGCCCAGCCCGGGCTGGCTGGCCTGCATGCGCCTCCTGTATAAGGCCCCA 779

FIG. 24A.

```
688  AGCCCG.CTGTCTCAGCCCTCCACTCCCTGCAGAGCTCAGAAGCGTGACC  736
     ||||| | ||||||||||||||||||||||||||||||||||||  ||||
780  AGCCTGCCTGTCTCAGCCCTCCACTCCCTGCAGAGCTCAGAAGCACGACC  829

737  CCAGCTGCAGCCATGAAGTGCCTCCTGCTTGC.....CCTGGCCCTCAC  780
     |||||||||||||||||||||||||||||||     |||||||||| |
830  CCAGCTGCAGCCATGAAGTGCCTCCTGCTTGCCCTGGGCCTGGCCCTCGC  879

781  CTGTGGCGCCCAGGCCCTCATCGTCACC  808
     |||||||| |||||||| ||||||||||
880  CTGTGGCGTCCAGGCCATCATCGTCACC  907
```

* TRANSLATION INITIATION CODON

FIG. 24B.

LINKER GP 278/279

TRANSLATION START SITE (αS_1 SIGNAL SEQUENCE)

ClaI

CGATAACCATGAAACTTCTTATCCTCACCTGTCTTGTGGCTGTTGCTCTTG

TATT — ETC.

αS_1 — hLZ SEQUENCE

BalI

—CCAAGGTCTTTGAAAGGTGTGAGTTGC

ETC.—AACC

FIG. 25.

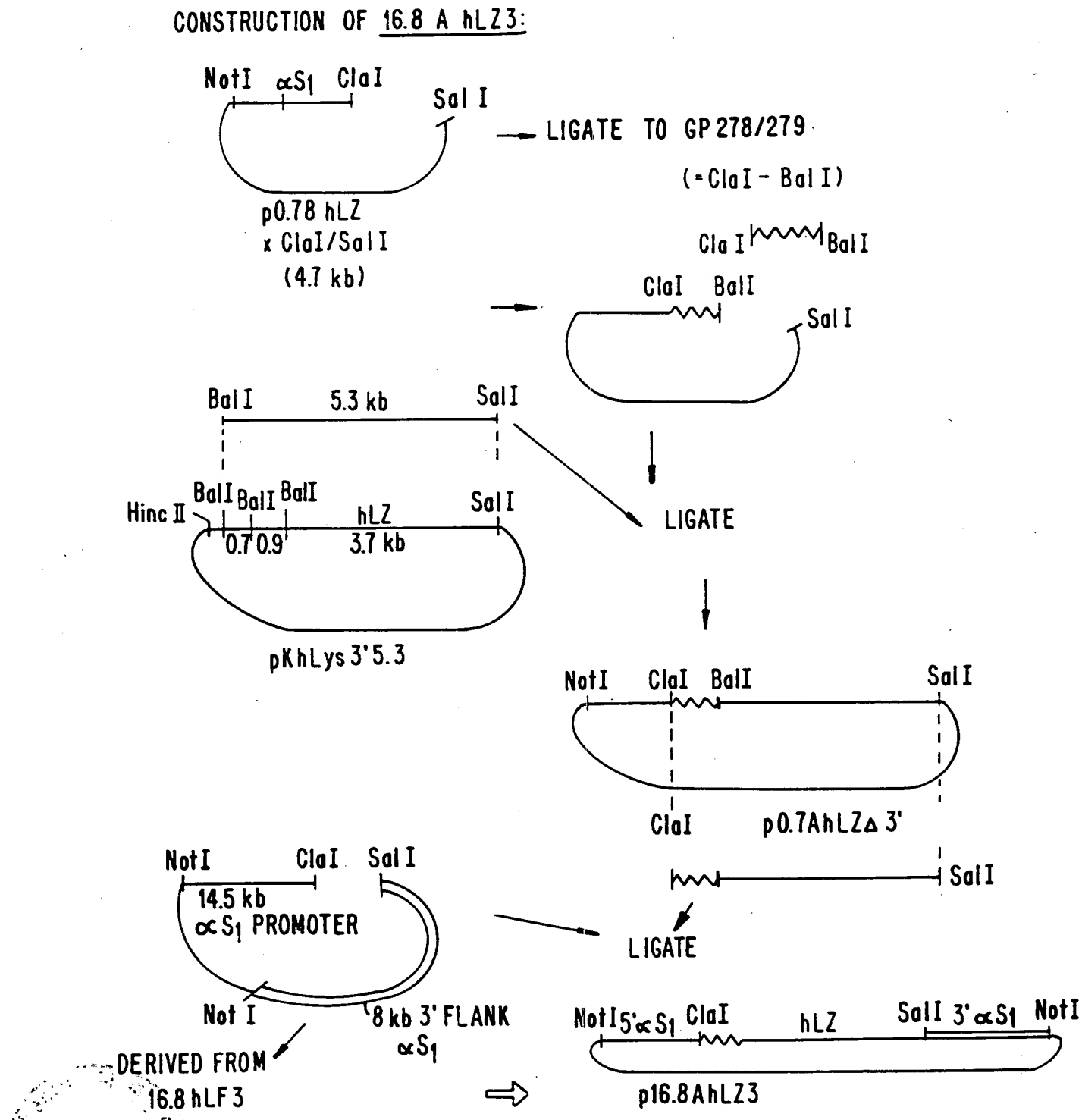

CONSTRUCTION OF 16.8 A hLZ3:
NotI
αS1
ClaI
Sal I
LIGATE TO GP 278/279
(=ClaI - Bal I)
p0.78 hLZ
x ClaI/SalI
(4.7 kb)
Cla I
BalI
ClaI
BalI
Sal I
BalI
5.3 kb
SalI
Hinc II
BalI
BalI
BalI
hLZ
SalI
0.7
0.9
3.7 kb
LIGATE
pKhLys 3'5.3
NotI
ClaI
BalI
SalI
ClaI
p0.7AhLZΔ 3'
SalI
NotI
ClaI
SalI
14.5 kb
αS1 PROMOTER
LIGATE
Not I
8 kb 3' FLANK
αS1
DERIVED FROM
16.8 hLF 3
NotI
5'αS1
ClaI
hLZ
SalI
3'αS1
NotI
p16.8AhLZ3

FIG. 26.

CONSTRUCTION OF 16 A hLZ3:

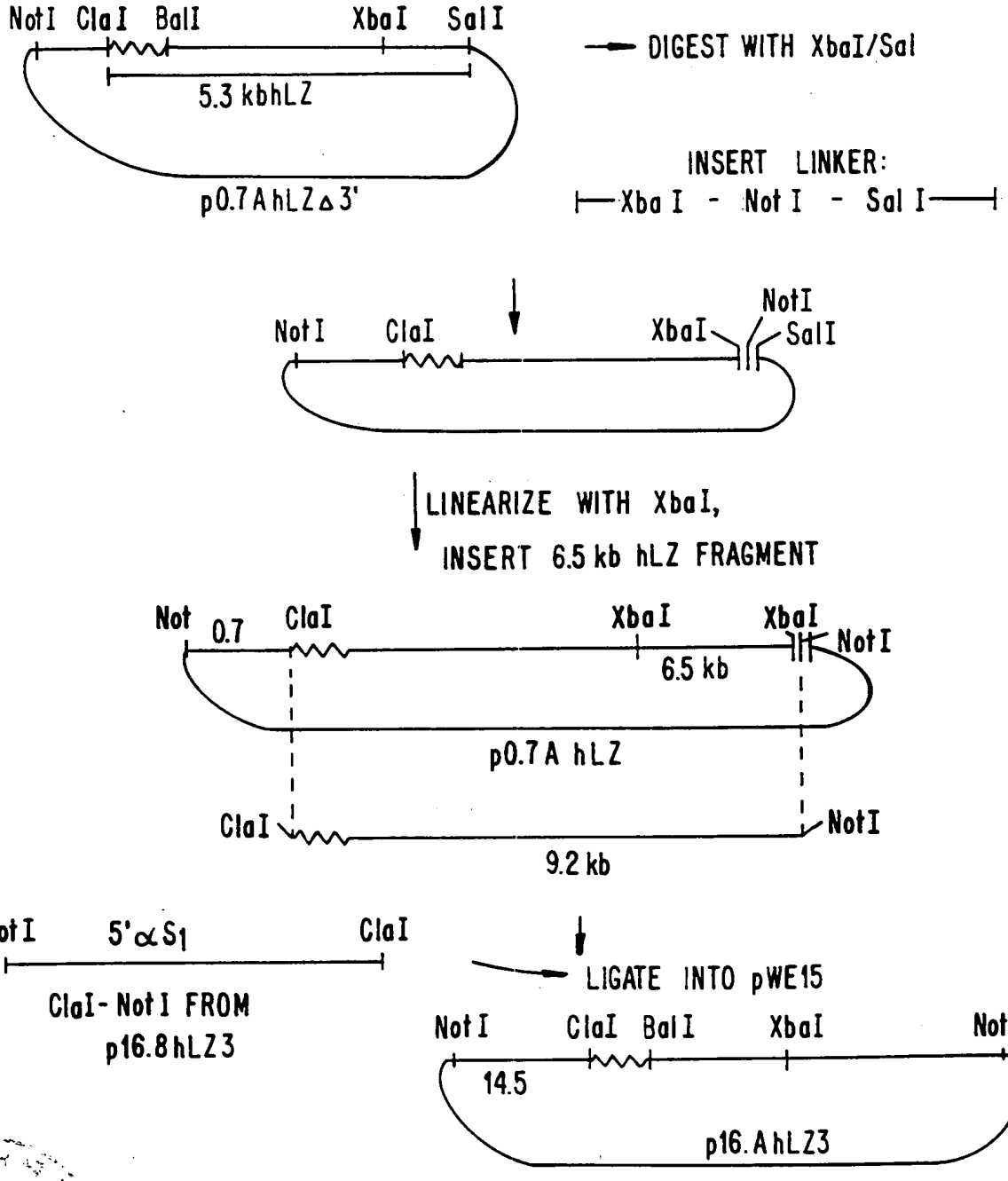


FIG. 27.